VNIVERSITAT DE VALÈNCIA

Departamento de Química Física

Instituto de Ciencia Molecular

Desarrollo de coordenadas del entorno para el estudio de la catálisis enzimática

Developing environmental coordinates for the study of enzymatic catalysis.

Rafael García Meseguer

Tesis Doctoral

Valencia, 2016

Ignacio Nilo Tuñón García de Vicuña, catedrático de Química Física del Departament de Química Física de la Universitat de València, José Javier Ruiz Pernía, investigador contratado del Departament de Química Física i Analítica de la Universitat Jaume I de Castellón y Damien Laage director de investigación del Chemistry Department de la École Normale Supérieure de París.

CERTIFICAN:

Que el trabajo con título: 'Desarrollo de coordenadas del entorno para el estudio de la catálisis enzimática' ha sido realizado por Don Rafael García Meseguer bajo nuestra dirección, para optar al grado de Doctor en Química.

Así, autorizan la presentación de este trabajo a efectos de seguir los trámites correspondientes de la Universitat de València.

Valencia, julio de 2016.

DIRECTORES

Dr. D. Ignacio Nilo Tuñón García de Vicuña

Dr. D. José Javier Ruíz Pernía

Dr. D. Damien Laage

A mi Pekeña.

Agradecimientos

Este trabajo pretende ser una recopilación del esfuerzo que a lo largo de cuatro años ha llevado una sola persona. Pero si algo puedo decir de esta tesis es que no la he hecho yo solo y me gustaría expresar mi agradecimiento a aquellas personas que han colaborado, en mayor o menor medida, en la culminación de esta tarea.

Antes de empezar con los agradecimientos personales, me gustaría agradecer al Ministerio de Educación, Cultura y Deporte por la financiación concedida; así como a la Universidad de Valencia por permitirme llevar a cabo mis estudios de doctorado.

En primer lugar, quiero agradecer a mi director Iñaki y a mis codirectores Javi y Damien, toda su colaboración y esfuerzo que me ha permitido sacar adelante esta tesis y me han ayudado a crecer como investigador. Gracias a ellos, ahora me siento preparado para dar el siguiente paso en mi carrera. También me gustaría agradecer a mi tía Margarita por ayudarme a que esta tesis luzca mucho mejor y con un inglés mucho más pulido.

Mi familia siempre ha sido una fuente de apoyo que me ha allanado el camino allá donde ha hecho falta para que yo pudiera seguir adelante, gracias por todo lo que me habéis dado.

Aunque día a día el trabajo puede ser una tarea solitaria, son los compañeros que están a tu alrededor los que la mejoran infinitamente y la hacen mucho más fácil. Es por eso que me gustaría agradeceros a vosotros: Kirill, Juan, Manuel, Maite, Juan Luis y Daryna; así como a todos los compañeros de Castellón, en especial a Isa, Kemel y Sergio; por vuestro papel en esta tesis que no es pequeño. También quiero hacer una mención especial a Julia que, yendo un paso por delante de mí, me dejabas miguitas de pan para que no me perdiera.

La Tesis no es un camino de rosas, mucho peor cuanto más al final estás, y la primera damnificada es tu vida social que se ve reducida considerablemente. Es por eso es que quiero agradecer a mis amigos, aquellos que han seguido ahí, por su paciencia, comprensión y apoyo. Tanto a los Warretes: Nacho, Luis, Ame, Emilio, Marta, Josevi, Santi, Bea y Aida; como a los Frikis: Baldur, Davis, Bolu, Fede, Salva, Owen, Luis e Iván. Habéis hecho que la parte de mi vida que no era La Tesis fuera mucho mejor. Por supuesto a mis compañeros de N7, por esos domingos pegando tiros que tanto me han ayudado a desconectar.

Tampoco quiero olvidarme de los nuevos amigos que han aparecido en estos últimos años, la mayoría durante mi estancia en el Colegio de España de París, donde he pasado tan buenos momentos a pesar de estar lejos de casa. A todos vosotros y en especial a: Evelin, Fer, Fuen, Anita, Vicente, Diego, Silviano, Ali, Gaizka, Breo y Ana; os doy las gracias pues habéis convertido el Colegio de España en mi segunda casa. También darles las gracias a Juan Pablo y a Gustavo, ya que juntos superar el Master fue mucho más fácil.

Por último, pero no por ello menos importante, mi más sincero agradecimiento quiero dedicárselo a Sandra, la persona que ha estado ahí a cada paso, que ha sufrido cada penuria conmigo, mi fuente de motivación y mi inspiración. Sin ti esta tesis no habría sido posible.

Index

Abbreviations

Symbols

Constants

k: Boltzmann's constant.

R: Ideal gas constant.

h: Planck's constant.

$\hbar$: Reduced Planck's constant.

Physical Magnitudes

k_r: Rate constant.

G: Gibbs' free energy.

V: Electrostatic Potential.

κ: Recrossing transmission coefficient.

γ: Tunneling coefficient.

W: Potential of mean force.

T: Temperature.

Coordinates

ξ: Reaction coordinate.

S: Solvent coordinate.

q: Antisymmetric transfer coordinate.

Q: Donor acceptor distance coordinate.

Ep: Electrostatic potential coordinate.

$\Delta\varepsilon$: Energy difference coordinate.

Units

Å: Angstrom.

$|e|$: Electron charge.

***fs*:** Femtosecond.

***ps*:** Picosecond.

***kcal*:** Kilocalories.

***kJ*:** Kilojoules.

***K*:** Kelvin.

mol: Mole.

***Deg*:** Degrees.

Acronyms

In alphabetic order.

COMT: Catechol O-Methyltransferase.

CTST: Conventional Transition State Theory.

DCE: Dichloroethane.

DFT: Density Functional Theory

DHF: Dihydrofolate.

DHFR: Dihydrofolate Reductase.

DhlA: Haloalkane Dehalogenase.

EFEP: Equilibrium Free Energy Path.

FES: Free Energy Surface.

FDH: Formate Dehydrogenase.

KIE: Kinetic Isotope Effect.

MD: Molecular Dynamics.

MM: Molecular Mechanics.

MFEP: Minimum Free Energy Path.

NAD: Nicotinamide Adenine Dinucleotide.

OPLS-AA: All-Atom Optimized Potential for Liquid Simulation.

PDB: Protein Data Bank.

PES: Potential Energy Surface.

PS: Product State.

QM: Quantum Mechanics.

RC: Reaction Coordinate.

RS: Reactant State.

SAM: S-adenosylmethionine

SC: Solvent Coordinate.

TS: Transition State.

TST: Transition State Theory.

VTST: Variational Transition State Theory.

ZPE: Zero Point Energy.

Chapter 1
Introduction

Enzymes are macromolecular biological catalysts that speed up most of the biological reactions in living organisms allowing them to take place in conditions and timescales compatible with life.[1] The atmospheric nitrogen fixation process is a good example of the catalytic power of enzymes. Being needed in a great number of biological processes, nitrogen is fully available in our atmosphere in the form of the diatomic molecule. In this form, nitrogen needs to be reduced to become biologically available, but its triple bond makes it a very stable molecule. In industry, this reaction requires an iron catalyzer, about 500 °C and 150-200 atmospheres to get around 10% of efficiency.[2] On the other hand, there are bacteria in nature able to perform the same process at room conditions of pressure and temperature thanks to an enzymatic complex named nitrogenase which reduces gas nitrogen to ammonia.[3] Enzymes also participate in many disease processes and usually are the

main target in the research of medical treatments, as is the case of the drug methotrexate, an inhibitor of the dihydrofolate reductase, an enzyme involved in the synthesis of nucleotides. The inhibition of this enzyme with methotrexate is often used in cancer chemotherapy,[4] as this drug is more toxic to rapidly growing cells, that have to carry out DNA replication more frequently, than to non-dividing cells.

But, where does the catalytic power of enzymes come from? Enzyme efficiency as catalyzers and their ability to carry on reactions in "soft" conditions have made them the target of many studies with the aim of answering this question. This topic, however, is still a matter of debate without a definite explanation.

Computer simulations have demonstrated to be a powerful ally in describing enzymatic processes and many different theoretical approaches have been proposed to address this task.[5-25] However, the first problem that we encounter when trying to model the enzymatic mechanism is the number of atoms involved in the process. While reactions in gas phase have been widely studied at high levels of Theory, enzymatic systems involve such a large number of particles that they are difficult to model, and calculations need to be heavily simplified. The need of a proper simulation of the environmental effects, together with the computational limitations, forces us to look for a compromise between the quality of our model and its computing availability. Hybrid QM/MM methodology is the most used model to approach this problem.[5,26-29] It allows the description of the atoms involved in the bond breaking and

forming process with quantum mechanical methods (QM) while the rest of the system is treated by means of classical molecular mechanics (MM).

The effects of the enzyme over the catalytic process can range over quite different phenomena and may play key roles in the enzymatic process. For instance, it is well known that many enzymatic processes are controlled by substrate binding or product release. In these cases, protein mobile loops can act as the gates to the active site, and then their motion can be the rate-determining step.[30] In other enzymes, the binding of the substrate can promote conformational changes in the enzyme that are needed to place some catalytic residues properly.[31] The environment created by the enzyme also plays an essential role in the chemical step of the reaction. This is, for example, the case of the enzyme formate dehydrogenase, which catalyzes the hydride transfer reaction from formate to NAD^+. When this process is modelled in water solvent, the charge neutralization that occurs during the process results in a high energy barrier.[32] On the other hand, in the enzyme a network of residues hydrogen-bonded to the reactants helps to place the formate in an adequate position for the reaction and facilitate the charge neutralization process.[32,33]

Regarding the effects of the enzyme on the chemical process, the question that puzzles many researchers is whether the catalytic ability of the enzyme plays its major role lowering the free energy barrier of the reaction or if, instead, the fluctuations and motions of the enzyme are directly affecting the rate constant of the reaction in a non-statistical way. The answer to this question does not seem to be an easy one. One of the

main problem lies, as I said before, in the complexity of the enzyme that makes difficult to define and quantify the effects of these fluctuations on the reaction rate constant. This idea that protein motions are coupled to the reaction –and hence contribute to catalysis–has been the object of many experimental and theoretical studies, both for and against this interpretation.[33-56] The development of new methodologies that were able to measure the evolution of the environment along the reaction could be the key to deepen in the role of protein motions.

That is precisely the objective of this thesis project: to develop, implement and apply two new methodologies that focus on the treatment of the environmental degrees of freedom and their coupling with the chemical reaction. In the first place, we will present the theoretical foundations needed to understand the developments and applications of this project, including some insight in the Transition State Theory, which will help us to have a better perspective of the problem of the so-called "dynamical effects". Secondly, we will provide and discuss the two methodological approaches developed and applied in this project. In chapter 4 we will show the results of the application of these two new methodological approaches to four different enzymatic systems: Haloalkane dehalogenase (DhlA), catechol-o methyltransferase (COMT), dihydrofolate reductase (DHFR) and formate dehydrogenase (FDH). Finally, chapter 5 will be a recapitulation of the conclusions taken from the results in chapter 4 and chapter 6 will include the papers published as a result of our research.

Bibliography

(1) Wolfenden, R. *Acc. Chem. Res.* **1972**, *5*, 10.

(2) Modak, J. M. *Reson* **2002**, *7*, 69.

(3) Lee, C. C.; R., M. W.; Hu, Y. In *The Metal-Driven Biogeochemistry of Gaseous Compounds in the Environment*; Kroneck, P. M. H., Torres, M. E. S., Eds.; Springer Netherlands: **2014**; Vol. 14, p 147.

(4) McGuire, J. J. *Curr. Pharm. Des.* **2003**, *9*, 2593.

(5) Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467.

(6) Warshel, A. *Proc. Nat. Acad. Sci. USA* **1978**, *75*, 5250.

(7) García-Viloca, M.; Alhambra, C.; Truhlar, D. G.; Gao, J. *J. Chem. Phys.* **2001**, *114*, 9953.

(8) Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W. *B. B. A.-Proteins Proteom.* **2006**, *1764*, 1647.

(9) Yadav, A.; Jackson, R. M.; Holbrook, J. J.; Warshel, A. *J. Am. Chem. Soc.* **1991**, *113*, 4800.

(10) García-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Science* **2004**, *303*, 186.

(11) Ruggiero, G. D.; Williams, I. H.; Roca, M.; Moliner, V.; Tuñón, I. *J. Am. Chem. Soc.* **2004**, *126*, 8634.

(12) Roca, M.; Andrés, J.; Moliner, V.; Tuñón, I.; Bertrán, J. *J. Am. Chem. Soc.* **2005**, *127*, 10648.

(13) Ruiz-Pernía, J. J.; Silla, E.; Tuñón, I.; Martí, S. *J. Phys. Chem. B* **2006**, *110*, 17663.

(14) Kanaan, N.; Roca, M.; Tuñón, I.; Martí, S.; Moliner, V. *J. Phys. Chem. B* **2010**, *114*, 13593.

(15) Tuñón, I.; Hynes, J. T. *Chem. Phys. Chem.* **2011**, *12*, 184.

(16) Zinovjev, K.; Ruiz-Pernía, J. J.; Tuñón, I. *J. Chem. Theo. Com.* **2013**, *9*, 3740.

(17) Duarte, F.; Amrein, B. A.; Blaha-Nelson, D.; Kamerlin, S. C. L. *Biochim. Biophys. Acta* **2015**, *1850*, 954.

(18) Mulholland, A. J. *Drug Discov. Today* **2005**, *10*, 1393.

(19) Antoniou, D.; Caratzoulas, S.; Kalyanaraman, C.; Mincer, J. S.; Schwartz, S. D. *Eur. J. Biochem.* **2002**, *269*, 3103.

(20) Antoniou, D.; Basner, J.; Núñez, S.; Schwartz, S. D. *Chem. Rev.* **2006**, *106*, 3170.

(21) Devi-Kesavan, L. S.; Gao, J. *J. Am. Chem. Soc.* **2003**, *125*, 1532.

(22) Pang, J.; Pu, J.; Gao, J.; Truhlar, D. G.; Allemann, R. K. *J. Am. Chem. Soc.* **2006**, *128*, 8015.

(23) Truhlar, D. G.; Gao, J.; García-Viloca, M.; Alhambra, C.; Corchado, J.; Luz Sánchez, M.; Poulsen, T. D. *Int. J. Quant. Chem.* **2004**, *100*, 1136.

(24) Pu, J.; Ma, S.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. B* **2005**, *109*, 8551.
(25) Kamerlin, S. C. L.; Warshel, A. *WIREs Comput. Mol. Sci.* **2011**, *1*, 30.
(26) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.
(27) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718.
(28) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700.
(29) Bash, P. A.; Field, M. J.; Davenport, R. C.; Petsko, G. A.; Ringe, D.; Karplus, M. *Biochemistry* **1991**, *30*, 5826.
(30) Clarke, A. R.; Wigley, D. B.; Chia, W. N.; Barstow, D.; Atkinson, T.; Holbrook, J. J. *Nature* **1986**, *324*, 699.
(31) O'Brien, J. R.; Schuller, D. J.; Yang, V. S.; Dillard, B. D.; Lanzilotta, W. N. *Biochemistry* **2003**, *42*, 5547.
(32) Castillo, R.; Oliva, M.; Martí, S.; Moliner, V. *J. Phys. Chem. B* **2008**, *112*, 10012.
(33) Roca, M.; Oliva, M.; Castillo, R.; Moliner, V.; Tuñón, I. *Chem. Eur. J.* **2010**, *16*, 11399.
(34) Villà, J.; Warshel, A. *J. Phys. Chem. B* **2001**, *105*, 7887.
(35) Kamerlin, S. C. L.; Warshel, A. *Proteins* **2010**, *78*, 1339.
(36) Warshel, A. *Proc. Nat. Acad. Sci. USA* **1984**, *81*, 444.
(37) Careri, G.; Fasella, P.; Gratton, E. *Ann. Rev.of Biophys. Bioeng.* **1979**, *8*, 69.
(38) Gavish, B.; Werber, M. M. *Biochemistry* **1979**, *18*, 1269.
(39) McCammon, J. A.; Wolynes, P. G.; Karplus, M. *Biochemistry* **1979**, *18*, 927.
(40) Karplus, M.; McCammon, J. A. *Ann. Rev. Biochem.* **1983**, *52*, 263.
(41) Cannon, W. R.; Singleton, S. F.; Benkovic, S. J. *Nat. Struct. Mol. Biol.* **1996**, *3*, 821.
(42) Neria, E.; Karplus, M. *Chem. Phys. Lett.* **1997**, *267*, 23.
(43) Miller, G. P.; Benkovic, S. J. *Biochemistry* **1998**, *37*, 6327.
(44) Závodszky, P.; Kardos, J.; Svingor, A.; Petsko, G. A. *Proc. Nat. Acad. Sci. USA* **1998**, *95*, 7406.
(45) Kohen, A.; Cannio, R.; Bartolucci, S.; Klinman, J. P. *Nature* **1999**, *399*, 496.
(46) Kohen, A.; Klinman, J. P. *Chem. Biol.* **1999**, *6*, R191.
(47) Xie, X. S.; Lu, H. P. *J. Biol. Chem.* **1999**, *274*, 15967.
(48) Radkiewicz, J. L.; Brooks, C. L. *J. Am. Chem. Soc.* **2000**, *122*, 225.
(49) Billeter, S. R.; Webb, S. P.; Agarwal, P. K.; Iordanov, T.; Hammes-Schiffer, S. *J. Am. Chem. Soc.* **2001**, *123*, 11262.
(50) Eisenmesser, E. Z.; Bosco, D. A.; Akke, M.; Kern, D. *Science* **2002**, *295*, 1520.
(51) Daniel, R. M.; Dunn, R. V.; Finney, J. L.; Smith, J. C. *Ann. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 69.

(52) Thorpe, I. F.; Brooks, C. L. *J. Phys. Chem. B* **2003**, *107*, 14042.

(53) Nam, K.; Prat-Resina, X.; García-Viloca, M.; Devi-Kesavan, L. S.; Gao, J. *J. Am. Chem. Soc.* **2004**, *126*, 1369.

(54) Bhabha, G.; Lee, J.; Ekiert, D. C.; Gam, J.; Wilson, I. A.; Dyson, H. J.; Benkovic, S. J.; Wright, P. E. *Science* **2011**, *332*, 234.

(55) Loveridge, E. J.; Behiry, E. M.; Guo, J.; Allemann, R. K. *Nat. Chem.* **2012**, *4*, 292.

(56) Tuñón, I.; Laage, D.; Hynes, J. T. *Arch. Biochem. Biophys.* **2015**, *582*, 42.

Chapter 2
Theoretical Foundations

Theoretical and Computational Chemistry has evolved intensively since the first quantum theories appeared. Many models and methodologies have been developed and computational tools became more popular and accessible.

If we want to improve and develop new methodologies, first we need to understand the pillars on which they are going to rest. In this chapter I will present and explain the main principles on which this Thesis is based.

2.1. Molecular Mechanics

Many studies published about enzymes make use of classical molecular mechanics methods (MM).[1-9] These methods describe the atoms of the system as point charges with Van der Waals parameters, to calculate the non-bonding interactions. Energy changes associated to bonds distances,

angles and dihedrals are represented by potential energy functions which depend on force constants and equilibrium values. In general, the total potential energy of the system can be expressed as:

$$E_{Tot} = E_{bonding} + E_{nonbonding} \quad (2.1)$$

$$E_{bonding} = E_{bonds} + E_{angles} + E_{dihedrals} \quad (2.2)$$

$$E_{nonbonding} = E_{electrostatic} + E_{Van\ der\ Waals} \quad (2.3)$$

The $E_{nonbonding}$ term is calculated only for those atom-pairs belonging to different molecules, or separated by at least 3 bonds.

These functions and parameters constitute a force field. Force fields differ in the functional forms selected to represent the different energy terms, the atoms types considerer and in the parameterization strategy employed. Some of the most used force fields are:

- Chemistry at Harvard Macromolecular Mechanics (CHARMM)[1] which is a widely used force field developed by Martin Karplus at 1983.

- Assisted Model Building with Energy Refinement (AMBER),[4] originally developed by Peter Kollman's group at the University of California, San Francisco.

- Groningen Molecular Simulation (GROMOS)[10] is a force field for molecular dynamics simulation developed at the University of Groningen

- All-Atom Optimized Potential for Liquid Simulation (OPLS-AA)[3,5,8] which was developed by Prof. William L. Jorgensen at Purdue University and Yale University.

Throughout this thesis, OPLS-AA has been the force field employed for all the calculations.

Boundary conditions

One of the advantages of using MM methodologies is the possibility to simulate the system in condensed phases, where the number of atoms can be really large. However, the size of the systems that can be simulated is still too small compared to a macroscopic one. In order to avoid problems associated to the size of the simulated system one has to add boundary conditions that mimic the effect of the surroundings.

There are several approximations available to model the environment of a system and probably the most widely employed are the Periodic Boundary Conditions (PBCs).[11] In this approximation an infinite system is constructed as a periodically repeated array of the finite system that is being studied (see Figure 2.1). Of course, for the infinite system to be continuous, the finite system must be of such a shape that, when replicated, can fill the space. The most common option in three dimensions is to use cubic orthorhombic boxes.

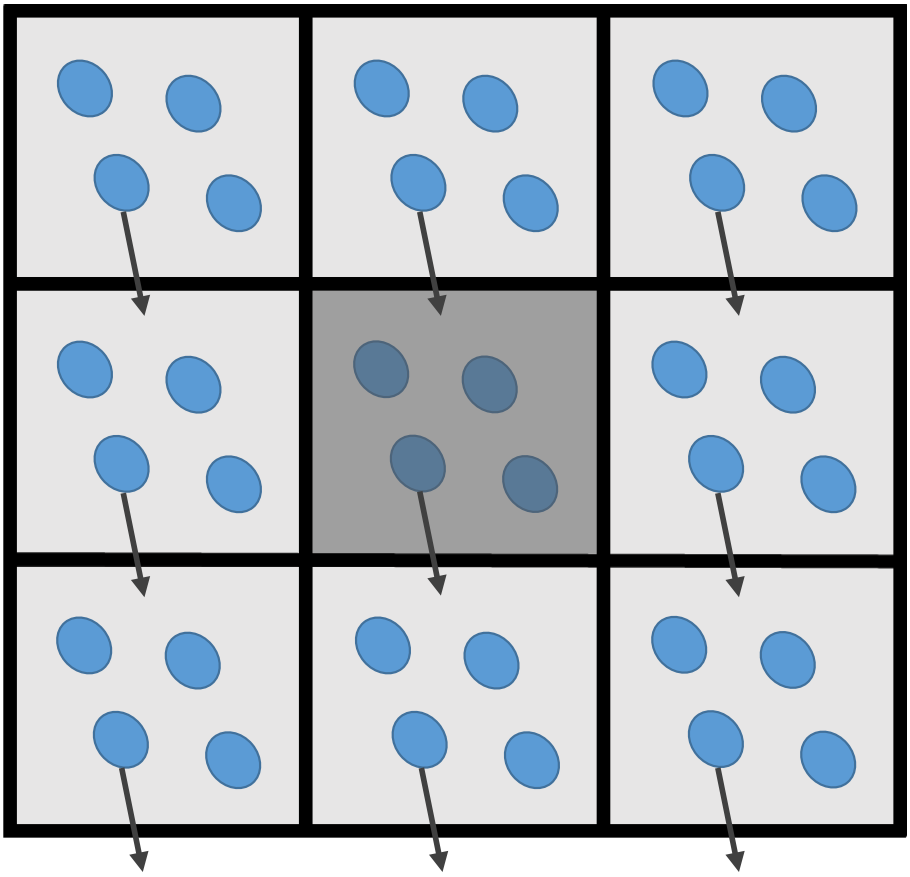


Figure 2.1: *Schematic representation of periodic boundary condition in two dimensions. The shaded box is replicated in both dimensions. A particle moving out of a box will be replaced by one moving in from the opposite side.*

This methodology is usually combined with the nearest neighbour approximation in such a way that a particle will only interact with the nearest copy of another particle in the system, i.e. it will not interact with a copy of itself and it will only interact with the particles of adjacent replicated systems.

Another strategy is the use of non-periodic boundary conditions, where spherical constraints are imposed, leaving free only those atoms contained within the sphere.[12] This separation between free and frozen atoms, can be combined with a buffer region where the motion of the particles is constrained. The computational cost of applying this methodology is much lower and it is very useful for very large systems where the simulation of the whole system becomes impracticable with periodic boundary conditions.

Long range interactions.

The calculation of $E_{nonbonding}$ runs over all pairs of interacting atoms in the system. Because the number of these pairs increases significantly with every atom added to the system, the calculation of the non-bonding energy rapidly becomes unmanageable. To overcome this problem and reduce the number of nonbonding interactions calculated, the interaction function is modified so that the interaction between atoms after a finite distance is set to zero. These are the cutoff methods and they can be:

- Truncated, where every interaction further than a determined value of distance is ignored.
- Smoothing, where the energy softly decays to zero in a given range.

While the truncated method is much simpler, it is certainly not the most precise as it introduces discontinuities to the energy function and its derivatives. The alternative, the smoothing method, usually makes use of two radius, r_{in} and r_{out} where every interaction with a distance larger

than r_{out} is zeroed and all interactions inside the r_{in} radius are fully calculated. For the distances in between, their energies of interaction are softened depending on the distance.[13]

The cutoff procedure is used not only for van der Waals interactions but also for electrostatic interactions. However, the use of truncation procedures is seriously questionable for long-range interactions, since the electrostatics decay very slow. The Ewald method is a well-known technique for calculating electrostatic interactions in crystals, as well as a standard method for simulations of dielectric properties of polar liquids.[14,15] In this method the long-range interaction is divided into two parts: a short-range contribution, and a long-range contribution. The short-range contribution is calculated in real space, whereas the long-range contribution is calculated using a Fourier transform. The advantage of this method is the rapid convergence of the energy compared with that of a direct summation.[15]

MM force fields can calculate, with fairly good accuracy, the potential energy of the system, even when it includes several thousands of particles. But classical methods cannot describe accurately bond cleavage and electron redistribution, only quantum mechanics methodologies are capable of such representations.

2.2. Quantum Mechanics

Quantum Mechanics (QM) calculations for molecular systems usually requires the resolution of Schrödinger equation in its time independent form:

$$\hat{H}|\Psi\rangle = E|\Psi\rangle \tag{2.4}$$

where Ψ is the wavefunction of the system and $\hat{H}$ is the Hamiltonian operator of the system.

$$\hat{H} = \hat{K}_\alpha + \hat{K}_i + \hat{U}_{\alpha\alpha} + \hat{U}_{ii} + \hat{U}_{i\alpha} \tag{2.5}$$

The Hamiltonian contains the kinetic energy of the nuclei ($\hat{K}_\alpha$) and of the electrons ($\hat{K}_i$), as well as the potential energy of interaction among nuclei ($\hat{U}_{\alpha\alpha}$), electrons ($\hat{U}_{ii}$) and among nuclei and electrons ($\hat{U}_{i\alpha}$). The interaction terms prevent this equation to be solved exactly. To obtain approximate solutions, the Born-Oppenheimer approximation[16] assumes that we can consider independently the movement of the electrons while the nuclei are stationary. Therefore, for each position of the nuclei we have a different approximate solution of the electronic part of the Schrödinger equation. From this approach the complete wavefunction can be divided into two wavefunctions. On the one hand we have an electronic wavefunction $\Psi^{e}(\boldsymbol{r};\boldsymbol{R})$ dependent on the coordinates of the electrons and parametrically on those of the nuclei. On the other hand, we have a nuclear wavefunction $\Psi^{n}(\boldsymbol{R})$, dependent solely on the nuclear coordinates.

$$\Psi(\boldsymbol{R},\boldsymbol{r}) = \Psi^{n}(\boldsymbol{R})\Psi^{e}(\boldsymbol{R},\boldsymbol{r}) \tag{2.6}$$

The electronic wavefunction can be solved from the electronic Hamiltonian:

$$\hat{H}_e = \hat{K}_i + \hat{U}_{ii} + \hat{U}_{i\alpha} \tag{2.7}$$

$$\hat{H}_e\Psi^{e}(\boldsymbol{r};\boldsymbol{R}) = E_e(\boldsymbol{R})\Psi^{e}(\boldsymbol{r};\boldsymbol{R}) \tag{2.8}$$

The potential energy of the system will be defined by the position of the nuclei by the following equation:

$$E_{Tot}(\boldsymbol{R}) = E_e(\boldsymbol{R}) + U_{\alpha\alpha}(\boldsymbol{R}) \tag{2.9}$$

Then the Hamiltonian of the system for each static position of the nuclei will be of the following form:

$$\hat{H}_{QM} = \hat{K}_i + \hat{U}_{ii} + \hat{U}_{i\alpha} + U_{\alpha\alpha} \tag{2.10}$$

The next step to solve the Schrödinger equation will be to obtain the nuclear wavefunction; but in many cases, as it has been done in this Thesis, nuclei are considered as classical particles. This approximation is valid for heavy atoms and cannot account for quantum effects such as tunneling.[17]

There are many different ways to solve the Schrödinger equation after applying the Born-Oppenheimer approximation and, if applied correctly,

they can provide accurate results for molecular systems. But the computational cost increases considerably with the number of electrons. Usually the larger the system, the less rigorous the method of choice will be. In a protein system there are thousands of electrons and nuclei making impossible for any QM method to solve its Schrödinger equation for all the configurations accessible to the system.[18,19]

2.3. Hybrid QM/MM Methods

In the 70's Michael Levitt, Arieh Warshel and Martin Karplus developed a hybrid approach combining the accuracy of the QM with the efficiency and viability of the MM approaches.[20-22] They won Nobel Prize in 2013 "for the development of multiscale models for complex chemical systems". In Hybrid QM/MM methodology the largest part of the system is described with classical MM potentials while the QM treatment is reserved for that part of the system where the chemical reaction takes place (see Figure 2.2).

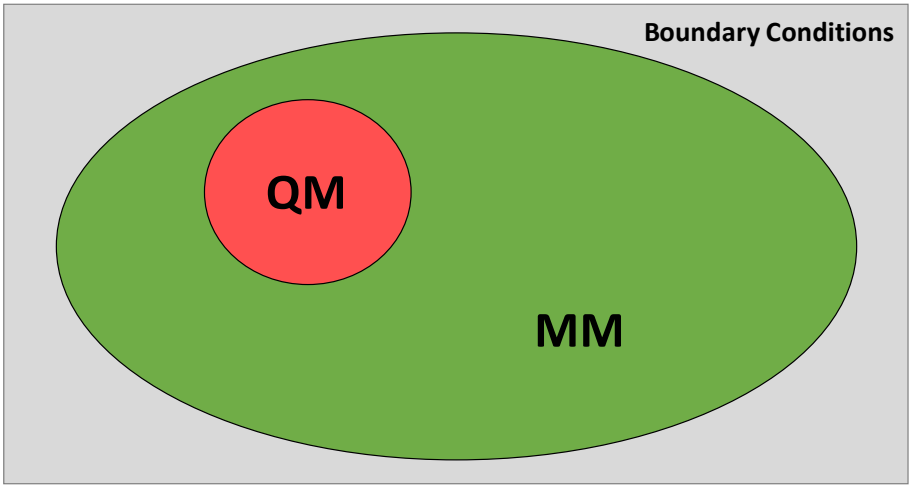


Figure 2.2: *Schematic representation of the division of the system in the QM/MM hybrid methodology.*

The Hamiltonian of the full system will be the summation of the Hamiltonians of the QM and MM parts in addition to the Hamiltonian interaction between the two subsystems.

$$\widehat{H}_{Total} = \widehat{H}_{QM} + \widehat{H}_{MM} + \widehat{H}_{QM/MM} \quad (2.11)$$

In the QM part, both electrons and nuclei are treated explicitly. On the other hand, the MM part contains the rest of the atoms, which are treated classically and the Hamiltonian of this part will be the total energy of the classical atoms.

$$\widehat{H}_{MM} = E_{Classical} = E_{bonding} + E_{nonbonding} \quad (2.12)$$

QM/MM Interaction

The last term in equation (2.11) describes interaction between the QM and MM parts. If the MM region is non-polarizable and the van der Waals energy term is described by means of a Lennard-Jones expression this Hamiltonian can be written as:[13]

$$\widehat{H}_{QM/MM} = -\sum_{iM} \frac{q_M}{r_{iM}} + \sum_{\alpha M} \frac{Z_\alpha q_M}{R_{\alpha M}} + \sum_{\alpha M} \left(\frac{A_{\alpha M}}{R_{\alpha M}^{12}} - \frac{B_{\alpha M}}{R_{\alpha M}^{6}} \right) \quad (2.13)$$

The i and α subscripts refer to the electrons and the nuclei of the QM part respectively; the M subscript, on the other hand, makes reference to the MM atoms. The first term represents the electrostatic interaction between the MM point charges (q_M), without any polarization, and the electrons of the QM part; the second one is the electrostatic interaction between the nuclei of the QM (Z_α)and MM part and the last one corresponds to the van der Waals interaction between both subsystems. The last two terms do not include electron coordinates, so they can be

calculated and directly added after solving the electronic structure. However, the first term does include electronic coordinates and thus it has to be included in the self-consistent procedure.

QM/MM Frontier

Ideally the QM part should include full molecules, and the division between the QM and MM part should not be set between two bonded atoms. However, in many cases there are residues of an enzyme directly participating in the reaction and must be treated by quantum mechanics or if one substrate is too big it can be divided between both regions to keep the QM part of a reasonable size. Therefore, the frontier between the two subsystems involve cutting a covalent bond and an approximation will be needed to fulfill the valence of the frontier quantum atom.

One of the most common methodologies is the link atom. This atom, which is usually defined as a hydrogen atom, is placed between the other two atoms affected by the frontier separation and it is treated quantum mechanically within the QM part.[23,24]

There are other methods for treating the QM/MM frontier, such as the Local Self Consistent Field (LSCF)[25-28] and the Generalized Hybrid Orbital (GHO).[28,29] In these methods, the frontier bonds are described by frozen bond orbitals, whose characteristics depend on the method used.

2.4. Potential Energy Landscapes

Using the hybrid QM/MM methodology we can calculate the energy of the system as a function of the coordinates of the nuclei, i.e. for each configuration of the nuclei we can obtain a value for the potential energy. As we change those coordinates we will obtain a Potential Energy (hyper) Surface (PES)

For the case of a diatomic molecule, the PES will be a function of only one parameter. For a linear system of three atoms there are two degrees of freedom (as long as the molecule stays linear) and the PES can be represented in 3 dimensions. But as the system gets more complicated, the number of variables increases and the representation becomes impossible. The number of calculations required to explore the PES for large systems is simply too high to become affordable.

Fortunately, not every configuration of the system is chemically viable, since most of them are so high in energy that the system cannot be found in that situation. We could then focus only in regions of the hypersurface that are chemically relevant. Of particular importance are the so called stationary structures (see Figure 2.3). To locate these stationary points, we need to know the gradient vector of the PES, as it will be zero for all of them, and the curvature of the energy surface around the zero-gradient structure. This curvature is stored in the second derivative of the Hessian matrix. Depending on the eigenvalues of the Hessian matrix, the stationary points can be of different kinds:

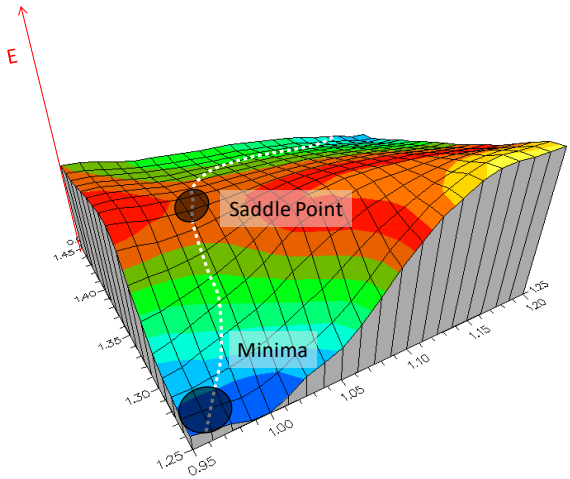


Figure 2.3: *Example of a PES in two dimensions with the stationary points marked with black circles.*

- Minima, with only positive eigenvalues, where any displacement from the structure will produce an increase in the energy of the system. These stationary points will include the reactant (RS), products (PS) and intermediate (IS) states of the reaction.

- Saddle points, with one negative eigenvalue, meaning that the displacement of the system in one direction will lead to a decrease in energy that will connect to the associated minima. This kind of stationary point is called a transition structure in the context of chemical reactions.

2.5. Molecular Dynamics

When we study a system in a condensed phase, the total number of degrees of freedom is huge. Such a PES presents a myriad of local minima which differ in small geometrical changes and have similar energies and properties (see Figure 2.4).[30]

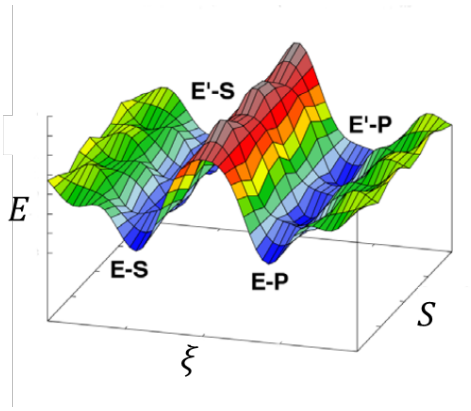


Figure 2.4: *PES for a chemical system condensed phase. The first coordinate (ξ) refers to the reacting system, while the second one (S) refers to the environment coordinates.*

The system can be found in different regions of the PES as long as they are energetically accessible. Stationary structures can be grouped in families and subfamilies which become accessible depending on the conditions of the experiment.[31] The properties of the system cannot be calculated from only one of the structures or families, but from an ensemble of them.

Molecular Dynamics (MD) is a widely used method to explore the different configurations and define the ensembles needed to understand the process under analysis.[24,32-41] It usually makes use of the equations of

motion of the particles in its classical form. Following Newton's equations to describe the movement:

$$\vec{\boldsymbol{F}}_{\boldsymbol{i}}(t) = m_i \vec{\boldsymbol{a}}_{\boldsymbol{i}} = m_i \frac{\delta \vec{\boldsymbol{v}}_{\boldsymbol{i}}}{\delta t} = m_i \frac{\delta^2 \vec{\boldsymbol{r}}_{\boldsymbol{i}}}{\delta t^2} \tag{2.14}$$

If we know the force ($\vec{\boldsymbol{F}}_i(t)$) that acts over a determined atom of mass m_i, as well as its position ($\vec{\boldsymbol{r}}_{\boldsymbol{i}}(t)$) at time t, it is possible to calculate the position it will have after a finite increment of time (Δt) by integrating the equation (2.14). One of the more efficient methods to perform such integration is the Velocity Verlet algorithm[42,43], in which the equations for the new position ($\vec{\boldsymbol{r}}_{\boldsymbol{i}}(t + \Delta t)$) and velocity ($\vec{\boldsymbol{v}}_{\boldsymbol{i}}(t + \Delta t)$) are:

$$\vec{\boldsymbol{r}}_{\boldsymbol{i}}(t + \Delta t) = \vec{\boldsymbol{r}}_{\boldsymbol{i}}(t) + \Delta t \cdot \vec{\boldsymbol{v}}_{\boldsymbol{i}}(t) + \frac{\Delta t^2}{2m_i} \cdot \vec{\boldsymbol{F}}_i(t) \tag{2.15}$$

$$\vec{\boldsymbol{v}}_{\boldsymbol{i}}(t + \Delta t) = \vec{\boldsymbol{v}}_{\boldsymbol{i}}(t) + \frac{\Delta t}{2m_i}\left(\vec{\boldsymbol{F}}_{\boldsymbol{i}}(t) + \vec{\boldsymbol{F}}_{\boldsymbol{i}}(t + \Delta t)\right) \tag{2.16}$$

To begin the dynamic simulation, a set of initial velocities for each atom must be included. Statistical thermodynamics tells us that the velocities of a classical system have a Maxwell-Boltzmann distribution and thus their initial values can be assigned initially form a gaussian distribution of velocities corresponding to the desired temperature.

Another important parameter to define is the time increment (Δt) that will be used in the simulation. It is considered as a general criterion that Δt has to be small enough so that the fastest vibration to be described can

be followed correctly. In practice for enzymatic systems Δt is usually equal to 1 femtosecond (fs) except when dealing with hydrogen transfer reactions, where Δt is 0.5 fs because of the faster vibration motion of the hydrogen atom.

Langevin Dynamics

In thermodynamics, a common definition of the state of a system; in which the number of particles (N), the volume of the system (V) and the temperature (T) are known; is that the statistical ensemble associated to this macroscopic description is the canonical or NVT ensemble. The NVT ensemble requires that the temperature is kept constant during the simulation. One of the methodologies to keep the temperature constant is the Langevin Dynamics (LD) method.[44,45]

The Langevin equation is a stochastic differential equation in which two force terms have been added to the Newton's second law to simulate the effects of the degrees of freedom not treated explicitly. The Langevin equation for the motion of atom i in the modified version of the velocity Verlet algorithm is:

$$m_i \cdot \vec{\boldsymbol{a}}_{\boldsymbol{i}}(t) = \vec{\boldsymbol{F}}(t) - \gamma_i m_i \vec{\boldsymbol{v}}_{\boldsymbol{i}}(t) + \vec{\boldsymbol{R}}_{\boldsymbol{i}}(t) \qquad (2.17)$$

where γ_i is the friction coefficient and $\vec{\boldsymbol{R}}_{\boldsymbol{i}}(t)$ is gaussian random force with zero mean and a variance of $\sqrt{2k_B T_0 \gamma_i m_i \delta(t-t')}$, being T_0 the thermostat temperature and $\delta(t-t')$ the Dirac's delta function. The term $\vec{\boldsymbol{R}}_{\boldsymbol{i}}(t)$ serves as a stochastic force responsible for random collisions of the molecular system with imaginary particles of the environment.

2.6. Free Energy

A macroscopic system in thermodynamic equilibrium, with a number N of particles in a volume V, will visit several accessible microscopic states ($\Psi_i(N,V)$) with associated energies ($E_i(N,V)$). The number of microscopic states with the same energy is known as degeneration ($\Omega_i(N,V,E_i)$). For a NVT ensemble the probability of finding the system in a determined microscopic state i can be expressed with the following equation:[46]

$$P_i(N,V,T) = \frac{exp\left(-\frac{E_i}{kT}\right)}{\sum_j exp\left(-\frac{E_j}{kT}\right)} = \frac{exp\left(-\frac{E_i}{kT}\right)}{Q(N,V,T)} \qquad (2.18)$$

being k the Boltzmann constant, T the temperature of the system and $Q(N,V,T)$ the canonic partition function, which can be expressed as the summation over the energy levels:

$$Q(N,V,T) = \sum_j exp\left(-\frac{E_j}{kT}\right) = \sum_E \Omega_i(N,V,E) exp\left(-\frac{E}{kT}\right) \qquad (2.19)$$

Here we have made the assumption that one can sum up the different energy levels, thus we are assuming that those levels are quantized. When we apply classical mechanics the energy becomes a continuum and the energy is not defined by quantum numbers, but by an ensemble of coordinates and momentum of every particle in the system ($\boldsymbol{r^N}, \boldsymbol{p^N}$). If we define $H\left(\boldsymbol{r^N}, \boldsymbol{p^N}\right)$ as the classic energy, the corresponding expression for the canonical partition function will be:

$$Q(N,V,T) = C \int \cdots \int exp\left(-\frac{H(\boldsymbol{r^N},\boldsymbol{p^N})}{kT}\right) d\boldsymbol{r^N} d\boldsymbol{p^N} \qquad (2.20)$$

$$d\boldsymbol{r^N} = \prod_{i=1}^{N} d\boldsymbol{r_i} \;\; ; \; d\boldsymbol{p^N} = \prod_{i=1}^{N} d\boldsymbol{p_i} \qquad (2.21)$$

$$C = (N! \cdot h^{3N})^{-1} \qquad (2.22)$$

where C is a constant that takes into account two quantum principle: (Pauli's exclusion and Heisenberg indetermination principles) and h is Planck's constant.

The probability of finding the system in a given microscopic state, with coordinates and momentum in the range of $(\boldsymbol{r^N} + d\boldsymbol{r^N}, \boldsymbol{p^N} + d\boldsymbol{p^N})$, is defined through the probability density expression:

$$\rho_{NVT}(r^N, p^N) = \frac{exp\left(-\frac{H(\boldsymbol{r^N},\boldsymbol{p^N})}{kT}\right)}{\int \cdots \int exp\left(-\frac{H(\boldsymbol{r^N},\boldsymbol{p^N})}{kT}\right) d\boldsymbol{r^N} d\boldsymbol{p^N}} \qquad (2.23)$$

Mechanical properties have values defined in their microscopic states. Therefore, their macroscopic values will be just the average over the corresponding microscopic values. Thus a mechanical property as the internal energy of the system (U) can be expressed as:

$$U = \int \cdots \int \rho_{NVT}\, H(\boldsymbol{r^N},\boldsymbol{p^N}) d\boldsymbol{r^N} d\boldsymbol{p^N} \qquad (2.24)$$

Thermal properties as entropy (S) and Helmholtz's free energy (A) depend also on how the system visit the different microscopic states. For example, the statistical thermodynamic expression for the Helmholtz's free energy is:

$$A = -kT \cdot ln\, Q(N, V, T) \tag{2.25}$$

It is not easy to use directly this equation to obtain a value for the free energy, but it is far easier to calculate the free energy difference (ΔA) between two states such as in the free energies of activation and reaction, which are relevant magnitudes in a chemical reaction.

It is important to keep in mind that we are discussing only the case for the Helmholtz's free energy obtained from the canonical ensemble. However, in condensed phase the difference between Gibbs' (G) and Helmholtz's free energy is negligible, because the volume does not change significantly, so we will use them indistinctly.

2.7. Potential of Mean Force

The free energy difference between two states I and II can be obtained from equation (2.25) as:

$$\Delta A = A_{II} - A_I = -kT \cdot ln\frac{Q_{II}}{Q_I} = \\ = -kTln\frac{\int\cdots\int exp\left(-\frac{H_{II}(\boldsymbol{r^N},\boldsymbol{p^N})}{kT}\right)d\boldsymbol{r^N}d\boldsymbol{p^N}}{\int\cdots\int exp\left(-\frac{H_I(\boldsymbol{r^N},\boldsymbol{p^N})}{kT}\right)d\boldsymbol{r^N}d\boldsymbol{p^N}} \tag{2.26}$$

There are several methods to calculate free energy differences.[47-51] Here I will focus in the use of the Potential of Mean Force (PMF) that provides the free energy change along a defined coordinate ξ and will be noted as $W(\xi)$.[52,53] Its value can be obtained evaluating the previous partition functions in equation (2.26) only for those configurations of the system that present particular values of the selected coordinate:

$$\Delta W(\xi_I \to \xi_{II}) = \\ = -kTln\frac{\int\cdots\int \delta\left(\xi\left(\boldsymbol{r^N}\right)-\xi_{II}\right)\cdot exp\left(-\frac{H(\boldsymbol{r^N},\boldsymbol{p^N})}{kT}\right)d\boldsymbol{r^N}d\boldsymbol{p^N}}{\int\cdots\int \delta(\xi(\boldsymbol{r^N})-\xi_I)exp\left(-\frac{H(\boldsymbol{r^N},\boldsymbol{p^N})}{kT}\right)d\boldsymbol{r^N}d\boldsymbol{p^N}} \tag{2.27}$$

The selection of configurations is driven by the Dirac's delta function ($\delta\left(\xi\left(\boldsymbol{r^N}\right)-\xi\right)$), which restricts the integral to those configurations with the desired value of the chosen coordinate.

If we consider the probability density function of finding a particular value of ξ in the system:

$$\rho(\xi) = \int \cdots \int \rho_{NVT}\,\delta(\xi(\boldsymbol{r^N}) - \xi)\, dr^N dp^N =$$

$$= \frac{\int \cdots \int \delta(\xi(\boldsymbol{r^N}) - \xi) \cdot exp\left(-\frac{H(\boldsymbol{r^N},\boldsymbol{p^N})}{kT}\right) d\boldsymbol{r^N} d\boldsymbol{p^N}}{\int \cdots \int exp\left(-\frac{H(\boldsymbol{r^N},\boldsymbol{p^N})}{kT}\right) d\boldsymbol{r^N} d\boldsymbol{p^N}} \tag{2.28}$$

The ratio between the probabilities of two different values of the coordinate ξ is:

$$\frac{\rho(\xi_{II})}{\rho(\xi_I)} =$$

$$= \frac{\int \cdots \int \delta(\xi(\boldsymbol{r^N}) - \xi_{II}) \cdot exp\left(-\frac{H(\boldsymbol{r^N},\boldsymbol{p^N})}{kT}\right) d\boldsymbol{r^N} d\boldsymbol{p^N}}{\int \cdots \int \delta(\xi(\boldsymbol{r^N}) - \xi_I) \cdot exp\left(-\frac{H(\boldsymbol{r^N},\boldsymbol{p^N})}{kT}\right) d\boldsymbol{r^N} d\boldsymbol{p^N}} \tag{2.29}$$

Substitute equation (2.29) in equation (2.27) to obtain:

$$\Delta W(\xi_I \to \xi_{II}) = -kT \cdot \ln \frac{\rho(\xi_{II})}{\rho(\xi_I)} \tag{2.30}$$

or alternatively:

$$W(\xi) = C' - kT \cdot \ln \rho(\xi) \tag{2.31}$$

where C' is an arbitrary constant.

This methodology allows us to calculate the PMF profile from the probability density of finding the system in a given value of the coordinate ξ. This probability density can be determined from the histogram corresponding to the distribution of the variable ξ:

$$\rho(\xi) \cdot \Delta\xi = \frac{\langle N(\xi) \rangle}{M} \tag{2.32}$$

being M the total number of configurations and $N(\xi)$ the number of configurations with the desired value of the coordinate.

Umbrella Sampling

This procedure requires that all the values used to calculate the PMF have a representative number of $N(\xi)$ structures. However, as seen in equation (2.28), the probability of finding the system in given value of the coordinate decays exponentially with its energy, which can lead to a poor sampling of the higher energy values of the coordinate ξ. Thus, to obtain converged values of the PMF we need to enhance the sampling of those configurations of interest whose energy is too high for the system to visit them spontaneously.

The umbrella sampling procedure[54] includes in the simulation an additional biasing potential $V_{umb}(\xi)$ that keeps the system close to a determined value of coordinate ξ so that we can have a good sampling around that value. The new energy function will be:

$$H_{umb} = H + V_{umb}(\xi) \tag{2.33}$$

and the probability distribution obtained with the biasing potential will be:

$$\rho(\xi)_{biased} = \\ = \frac{\int \cdots \int \delta(\xi(r^N) - \xi) \cdot exp\left(-\frac{H+V_{umb}(\xi)}{kT}\right) d\boldsymbol{r^N} d\boldsymbol{p^N}}{\int \cdots \int exp\left(-\frac{H+V_{umb}(\xi)}{kT}\right) d\boldsymbol{r^N} d\boldsymbol{p^N}} \quad (2.34)$$

This biased probability can be related to that of the unbiased system:

$$\rho(\xi) = \rho(\xi)_{biased} \cdot \frac{\left\langle exp\left(-\frac{V_{umb}(\xi)}{kT}\right)\right\rangle}{exp\left(-\frac{V_{umb}(\xi)}{kT}\right)} \quad (2.35)$$

where the averages have been taken in the NVT ensemble.

The umbrella potential can be a simply harmonic function. If we want to keep the system around a particular value of ξ_{ref} the biasing potential then will be:

$$V_{umb}(\xi) = \frac{1}{2} K_{umb}\left(\xi - \xi_{ref}\right)^2 \quad (2.36)$$

where K_{umb} is the force constant used to keep the system at the reference value. For each reference value we have to apply a different umbrella potential (see Figure 2.5). The ensemble of configurations reached inside this umbrella potential is called a simulation window and while performing the simulations we have to ensure that configurations from one window overlap with the configurations of the adjacent one, so that every value along the ξ coordinate is adequately sampled.

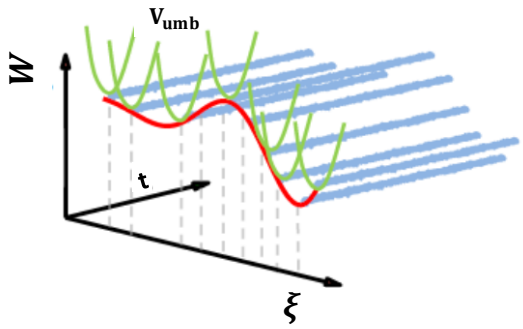


Figure 2.5: *Schematic representation of the calculation of a PMF. Each of the windows has been represented in blue with the umbrella potential applied depicted in green.*

Weighed Histogram Analysis Method

The Weighed Histogram Analysis Method (WHAM)[55] provides a procedure to combine the data obtained from the simulation using the umbrella sampling methodology to construct the PMF profile. According to this methodology the total distribution function is the weighed summation of the distribution function of the N_W windows calculated.

$$\rho(\xi) = \sum_{\alpha=1}^{N_W} w^{\alpha} \rho(\xi)^{\alpha} \tag{2.37}$$

and substituting eq. (2.35):

$$\rho(\xi) = \sum_{\alpha=1}^{N_W} w^{\alpha} \rho(\xi)^{\alpha}_{biased} exp\left(-\frac{V^{\alpha}_{umb} - F^{\alpha}}{kT}\right) \tag{2.38}$$

The weights included fulfill the normalization condition:

$$\sum_{\alpha=1}^{N_W} w^{\alpha} = 1 \tag{2.39}$$

and are introduced with the condition to minimize the statistical error (σ^2):

$$\frac{\partial \sigma^2[\rho(\xi)]}{\partial w^{\alpha}} = 0 \tag{2.40}$$

It can be demonstrated that the weights that fulfill these conditions are those with the form:[55]

$$w^{\alpha} = \frac{n^{\alpha} \cdot exp\left(-\frac{V_{umb}^{\alpha}(\xi) - F^{\alpha}}{kT}\right)}{\sum_{\beta=1}^{N_W} n^{\beta} \cdot exp\left(-\frac{V_{umb}^{\beta}(\xi) - F^{\beta}}{kT}\right)} \tag{2.41}$$

being n^{α} the number of independent points used to create the distribution function of the particular window α, and F^{α} a constant that needs to be estimated for that window α:

$$\begin{aligned} exp\left(-\frac{F^{\alpha}}{kT}\right) &= \left\langle exp\left(-\frac{V_{umb}(\xi)}{kT}\right)\right\rangle = \\ &= \int \cdots \int exp\left(-\frac{V_{umb}(\xi)}{kT}\right)\rho(\xi)d\xi \end{aligned} \tag{2.42}$$

As the constant F^{α} and the distribution function $\rho(\xi)$ are initially unknown, the problem is solved with an iterative process, where guessed

values are employed to obtain $\rho(\xi)$ and then the distribution function is used to calculate a new value of F^{α}. The process is carried out until convergence is reached.[13,56]

Free Energy Surfaces

The PMF concept can be generalized to analyze the free energy change as a function of more than one coordinates. For two coordinates named ξ and s the associated free energy surface can be obtained as:

$$W(\xi, s) = \\ = C' - kTln \int \rho(\boldsymbol{r^N})\, \delta(\xi(\boldsymbol{r^N}) - \xi_0)\delta(s(\boldsymbol{r^N}) - s_0) d\boldsymbol{r^N} \tag{2.43}$$

where $\rho(\boldsymbol{r^N})$ is the probability density of finding the system at configuration $\boldsymbol{r^N}$.

Free energy surfaces (FESs) allow a better understanding of reaction mechanisms in complex systems since we will be able to analyze the role and timing of each coordinate. This can be of particular interest in this Thesis if each coordinate can be associated to the changes in the chemical system and the environment, respectively.

2.8. Transition State Theory

The information obtained from MD simulations, in theory, should be enough to calculate any kinetic property including the rate constant of a chemical transformation (k_r). However, the amount of simulation time required to obtain directly this information with adequate accuracy makes this possibility unbearable and unpractical. However, as will be shown here, rate theories can be used to obtain expressions for the rate constant that can be evaluated with the information from biased MD simulations.

The first theoretical description of the dependence of the rate constant of a chemical reaction on the temperature was formulated by Svante Arrhenius,[57] who in 1889, based in experimental results and following the work of Jacobus Henricus van 't Hoff, provided and empirical relationship between the temperature and the rate constant in an equation bearing his name.

$$k_r = A \cdot e^{-E_a/RT} \tag{2.44}$$

where A is a pre-exponential factor or prefactor and E_a is the activation energy, R is the gas constant and T is the temperature. This equation derives from empirical observation and completely ignores the mechanism of the reaction. Thus, the source of the prefactor and activation energy parameters was rather obscure and further development was necessary.

In the 30s Eugene Wigner[58], Henry Eyring,[59] and Meredith Gwynne Evans with Michael Polanyi[60] developed simultaneously what was referred to as "activated-complex theory," "absolute-rate theory," and "theory of absolute reaction rates". All these theories are now known as Transition State Theory (TST). TST has undergone many improvements, becoming one of the most used theories in computational chemistry[61] because it provides an expression of the rate constant from molecular magnitudes that can be calculated theoretically. Nowadays TST stands as a general name used to describe any theory based in whole or in part on the following fundamental assumptions:[61,62]

- There is a hypersurface in phase space that divides it into reactant and product regions. Those structures that are on the vicinity of this hypersurface conform the Transition State (TS) ensemble.
- Trajectories passing through this hypersurface in the products direction and originated in the reactant region will not reach the hypersurface again before being stabilized in the product state.
- The reactant and transition states keep an equilibrium distribution.
- Motions along the reaction coordinate are separable from the remaining degrees of freedom and can be treated classically.

Reaction Coordinates

A Reaction Coordinate (RC) is an abstract one-dimensional parameter that represents the evolution of the system from reactants to products. In order to make use of a RC in our simulations, we need to define it as a

function of a set of coordinates of the system. The definition of the RC is not a trivial decision, as it will determine to a large extent the applicability of TST to a particular problem, including the representability of the TS ensemble and the free energy profile. The larger the system, the more degrees of freedom may contribute to the reaction and the definition of the RC to be used in TST might become more difficult.

Usually, RCs are defined using geometric valence coordinates directly involved in the reaction process. For example, if we want to study a dissociation reaction, the best choice for a distinguished RC would be the distance between the two dissociating atoms. Another very common process is transfer reactions, where an atom or a group of atoms are transferred from the donor (D), to the acceptor (A). A good RC definition can be the antisymmetric combination of those two distances:

$$\xi = d(D,T) - d(A,T) \tag{2.45}$$

At the reactants region, where the distance between the donor and the transferred atoms ($d(D,T)$) is larger than the distance between acceptor and the transferred atoms ($d(A,T)$), ξ will have negative values. In the case of a symmetric reaction, where the donor and acceptor molecules are equal, the saddle point of the reaction will be exactly at the same distances both for $d(D,T)$ and $d(A,T)$ so ξ will be equal to zero. However, in most cases where the reaction is not symmetric, the saddle point is usually close to $\xi = 0$. And finally on the products region, we have the opposite situation to the reactants, and thus the value of ξ will be positive.

For a condensed phase system, there is a huge number of degrees of freedom that may influence the reaction process in greater or lesser extent. How can we define the reaction coordinate then? In most cases we can still use geometric valence coordinates involving only the reacting system.[63-70] This is an easy definition for the RC, as it only requires a basic knowledge of the reaction mechanism and is the most common procedure. Unfortunately, this kind of definitions of the RC can be limiting, mainly because we lose information regarding the solvent coordinates. Moreover, in TST, the rest of degrees of freedom are considered to be in equilibrium at each value of the RC.[71] Deviations of this equilibrium assumption will cause that some trajectories reaching the dividing surface from reactants will end again in the reactant state. This effect can be corrected with the inclusion of a transmission coefficient (κ), which will be discussed in the following section. However, an alternative would be the inclusion of additional degrees of freedom in the definition of the RC. This definition of the RC will keep the environment in an adequate configuration at the TS, so the system can cross the dividing surface and reach the product state with less recrossing.[41] However, regardless of how many degrees of freedom are included in the RC, there will always be some amount of recrossing.[72,73]

In the case of reactions including the transfer of light particles this problem becomes even more significant because the geometric valence coordinates are faster than the environmental degrees of freedom, so the equilibrium assumption is less adequate. The RC must be better defined in terms of the environment and then assumed that the light particles can adapt adiabatically to it.[74-79]

Conventional Transition State Theory

Within this context we can identify the Conventional Transition State Theory (CTST) in which the dividing surface is placed at the saddle point on the PES and the RC is the normal mode associated to a negative force constant. Then the rate constant of the reaction is calculated as the one-way equilibrium flux of the structures reaching the saddle point from the reactants.[61,80]

If we take the equilibrium constant ($K^{\ddagger}$) in a bimolecular reaction between the RS ($[A][B]$) and the TS ($[TS]^{\ddagger}$):

$$A + B \rightarrow TS^{\ddagger} \rightarrow C \tag{2.46}$$

$$K^{\ddagger} = \frac{[TS]^{\ddagger}}{[A][B]} = \frac{Q_v^{\ddagger}}{Q_{v,A}Q_{v,B}} e^{-\frac{\Delta E_0}{kT}} \tag{2.47}$$

$$[TS]^{\ddagger} = \frac{Q_v^{\ddagger}}{Q_{v,A}Q_{v,B}} e^{-\frac{\Delta E_0}{kT}} [A][B] \tag{2.48}$$

The symbol ‡ makes reference to any variable associated with the TS. $Q_{v,A}$, $Q_{v,B}$ and $Q_v^{\ddagger}$ are the partition functions per volume unit of the RS and the TS respectively. ΔE_0 is the energy difference between the lowest energy level of the TS and the lowest energy level of the RS.

As we are only interested in the reactant to products flux rate we must account only half the number of $[TS]^{\ddagger}$. If the reaction is in equilibrium there will be the same number of $[TS]^{\ddagger}$ coming from reactants ($[TS]^{\ddagger}_{R\rightarrow P}$)

and from products ($[TS]^{\ddagger}_{P\to R}$). And because of the non-recrossing assumption, the concentration of the products will not affect the concentration of $[TS]^{\ddagger}_{R\to P}$

$$[TS]^{\ddagger} = [TS]^{\ddagger}_{R\to P} + [TS]^{\ddagger}_{P\to R} \tag{2.49}$$

$$[TS]^{\ddagger}_{R\to P} = [TS]^{\ddagger}/2 \tag{2.50}$$

$$\frac{dTS^{\ddagger}}{dt} = \frac{[TS]^{\ddagger}_{R\to P}}{t^{\ddagger}} = \frac{[TS]^{\ddagger}}{2t^{\ddagger}} \tag{2.51}$$

where $t^{\ddagger}$ is the average time the system takes to cross the dividing surface. As the system is in equilibrium and considering the motion along the RC as a classical translation in a box of size δ, $t^{\ddagger}$ can be expressed as:

$$t^{\ddagger} = \delta \cdot \left(\frac{\pi\mu}{2kT}\right)^{\frac{1}{2}} \tag{2.52}$$

where μ is the reduced mass corresponding to the movement along the RC. The separability assumption leads to

$$Q^{\ddagger} = Q^{\ddagger}_{v,RC}\bar{Q}^{\ddagger}_{v} \tag{2.53}$$

being $Q^{\ddagger}_{v,RC}$ the partition function corresponding to the motion along the RC which crosses the dividing surface of side δ, and $\bar{Q}^{\ddagger}$ the partition function of the remaining degrees of freedom at the transition state.

Considering again the motion along the RC as a classical translation:

$$Q_{v,RC}^{\ddagger} = \frac{(2\pi\mu kT)^{\frac{1}{2}}}{h} \cdot \delta \tag{2.54}$$

From equation (2.47) and (2.50):

$$\frac{dT^{\ddagger}}{dt} = \frac{kT}{h} \frac{\bar{Q}_v^{\ddagger}}{Q_{v,A} Q_{v,B}} e^{-\frac{\Delta E_0}{kT}} [A][B] \tag{2.55}$$

The experimental equation of the reaction rate for a bimolecular process is:

$$\frac{dT^{\ddagger}}{dt} = k_r(T)[A][B] \tag{2.56}$$

and then comparing (2.54) and (2.55):

$$k_r(T) = \frac{kT}{h} \frac{\bar{Q}_v^{\ddagger}}{Q_{v,A} Q_{v,B}} e^{-\frac{\Delta E_0}{kT}} \tag{2.57}$$

This equation can be expressed as a function of the equilibrium constant that excludes the contribution of the RC to the partition function of the TS.

$$k_r(T) = \frac{kT}{h} \bar{K}^{\ddagger} \tag{2.58}$$

$\bar{K}^{\ddagger}$ can be in turn expressed in terms of the activation free energy ($\Delta G_0^{\ddagger}$):

$$\Delta G_0^{\ddagger} = -RTln\overline{K}^{\ddagger} \tag{2.59}$$

and then:

$$k_r(T) = \frac{k_B T}{h} e^{-\frac{\Delta G_0^{\ddagger}}{RT}} \tag{2.60}$$

This is the equation for the rate constant derived from the CTST also known as the Eyring-Polanyi equation.[81,82]

Variational Transition State Theory

CTST is concerned basically with the properties of the saddle point, because knowledge of the PES in the surroundings of the saddle point and the reactant minimum is the unique information required. However, the assumption that the dynamical bottleneck of the reaction is in the saddle point is not completely accurate. Naturally, the saddle point is the highest potential energy point in the reaction path, but the rate constant depends on the free energy of activation, which includes entropic contributions. This fact leads to an underestimation of the reaction free energy and an overestimation of the rate constant.

Variational Transition State Theory (VTST)[83,84] is a generalization of the CTST that removes the restriction on the dividing surface to necessarily pass through the saddle point. In VTST the TS is defined as the point along the reaction path that maximizes the free energy and thus reduces the overestimation of the rate constant. It is important to note that a better description of the RC will take to account relevant degrees of freedom that reveal energy barriers not reflected in a poorly described one. Thus,

improving the definition of the RC will increase the free energy barrier and will result in a better calculation of the rate constant.

Grote-Hynes Theory

TST is a powerful tool for the calculation and analysis of the rate constant in a variety of gas and condensed-phase systems such as liquids, solids and, as in our case, enzymatic reactions. Even though, the accurate prediction of rate constants in condensed phases still presents a major challenge because of the number of degrees of freedom that participate in the process and the impossibility of reproducing all the interaction energies accurately. For simplicity we will define as "environment" those atoms that do not participate directly in the reaction. They usually belong to the MM section in a hybrid QM/MM calculation. On the other hand, "solute" and/or "substrate" will make reference to the atoms and molecules involved directly in the RC.

TST is based on the equilibrium assumption. Equilibrium solvation provides a good starting point for treating the effective force field created by the surrounding condensed-phase molecules and affecting the reacting molecules. This averaged potential is obtained from an equilibrium ensemble of configurations of solvent coordinates. Since this mean-field potential is obtained from an equilibrium ensemble averaged for a determined value of the reaction coordinate, the equilibrium assumption implies that the environment instantaneously equilibrates to every new value of the reaction coordinate. Equilibrium solvation neglects any dynamical influence from the solvent over the reaction dynamics that can appear from its fluctuations around equilibrium. Both

equilibrium and nonequilibrium solvent effects can induce recrossing of the TS surface, which leads to a breakdown of the fundamental assumption of TST.[85]

One solution to the non-equilibrium solvation problem is the inclusion of the transmission coefficient ($\kappa(T,\xi)$), dependent on the reaction coordinate (ξ), that takes into account the recrossing caused by environmental effects and thus reduces the rate constant accordingly with this effect. Grote and Hynes[86] derived an expression for the transmission coefficient where the reaction is studied through a simplified model. There, the reacting solute is treated as a single reaction coordinate while the rest of the system is modeled as a bath in terms of a generalized Langevin equation motion.[87-89] The Grote-Hynes expression may be interpreted as the effect of environmental friction on the motion across the barrier. Solvent friction and recrossing are just two different ways of looking at the same physical effect. The Grote-Hynes transmission coefficient can be obtained as the ratio between the equilibrium frequency (ω_{eq}), obtained from the free energy profile along the reaction coordinate, and the reactive frequency (ω_r), which can be obtained from the following relationship: [66,86]

$$\omega_r^2 - \omega_{eq}^2 + \omega_r \int_0^\infty \zeta_{TS}(t) \cdot e^{-\omega_r t} dt = 0 \qquad (2.61)$$

where $\zeta_{TS}(t)$ is the frictional kernel obtained from the autocorrelation function of the forces exerted on the reaction coordinate when the system is constrained at the transition state:[66]

$$\zeta(t) = \frac{\langle F_{RC}(0) F_{RC}(t) \rangle}{\mu_{RC} k_B T} \tag{2.62}$$

being $F_{RC}(t)$ the force of the reaction coordinate and μ_{RC} its reduced mass. The transmission coefficient will be of the form:

$$\kappa(T, \xi) = \frac{\omega_r}{\omega_{eq}} \tag{2.63}$$

TST for light particles

Transition state theory is basically a classical theory and therefore any quantum effect is completely neglected. This is due to the classical fundamental assumption that one can define the phase space surface dividing reactants and products and calculate the one-way flux through it. Specifying a one-way flux along the reaction coordinate and through the TS ensemble requires a precise value of the reaction coordinate and the sign of its conjugate momentum, but when one is specified precisely any knowledge of the other is forbidden by the uncertainty principle. Indeed, the uncertainty principle forbids the precise knowledge of any pair of non-commuting variables,[90] and all attempts to translate the one-way flux concept from the TST to the quantum language result in non-commuting variables[91] and thus uncertainty.

Tunneling Contribution

The inclusion of quantum effects that lead to quantized energy levels, Zero Point Energies (ZPE) and tunneling effects are essential for obtaining reliable values of the rate constant. The standard approach for

including quantum mechanical effects to account for the motion along the RC in TST goes through the introduction of a correction factor. Therefore, to correct the influence of tunneling on the rate constant we multiply the rate constant of the reaction by a ground-state transmission coefficient $\gamma(T)$.[92]

$$k_r(T) = \gamma(T)\frac{kT}{h}e^{-\frac{\Delta G_0^{\ddagger}}{RT}} \tag{2.64}$$

The tunneling transmission coefficient ($\gamma(T)$) is evaluated as the ratio between the semiclassical adiabatic ground state (SAG) probability and the quasiclassical probability:[93]

$$\gamma(T) = \frac{\int_0^{\infty} dE \cdot P^{SAG}(E) \cdot e^{-\frac{E}{kT}}}{\int_{V^{AG}}^{\infty} dE \cdot P^{C}(E) \cdot e^{-\frac{E}{kT}}} \tag{2.65}$$

where $P^{SAG}(E)$ and $P^{C}(E)$ are the semiclassical and classical probability respectively. $P^{C}(E)$ equals zero below V^{AG} (the maximum of the vibrationally adiabatic potential) and unity otherwise, so the transmission coefficient can be written as:

$$\gamma(T) = \frac{e^{\frac{V^{AG}}{kT}}}{kT}\int_0^{\infty} dE \cdot P^{SAG}(E) \cdot e^{-\frac{E}{kT}} \tag{2.66}$$

The tunneling transmission coefficient is evaluated by using an effective potential that will depend on the following approximations:

- Zero-curvature tunneling (ZCT),[94] which neglects the coupling between the reaction coordinate and the transverse modes.

- Small-curvature tunneling (SCT),[95-98] which incorporates such coupling, but considering that the curvature of MEP is not too large.

- Large curvature tunneling (LCT),[98-101] which considers straight trajectories as the tunneling paths and was especially designed for reactions with large curvature of the MEP and therefore with important quantum effects.

Ensemble-Averaged Variational Transition State Theory

Deviations from classical TST as a result of quantum tunneling effects in enzymatic reactions can be included in the free energy calculations by means of the Ensemble-Averaged Variational Transition State Theory (EA-VTST).[102-104] In this approach, the rate constant is expressed as:

$$k_r(T) = \gamma(T)\frac{k_B T}{h} e^{-\frac{\Delta G_{ac}^{QC}(T,\xi)}{RT}} \qquad (2.67)$$

ΔG_{ac}^{QC} is the quasi-classical activation free energy at the TS, obtained from the PMF ($W(T,\xi)$) and including a correction for quantizing the vibrations orthogonal to the reaction coordinate and the vibrational free energy of the reactant mode that correlates with motion along the reaction coordinate, and is calculated as:

$$\Delta G_{ac}^{QC} = [W(T,\xi^{\ddagger}) + \Delta W_{vib}(T,\xi^{\ddagger})] - \left[W(T,\xi_{RS}) + \Delta W_{vib,RS}(T) + G_{RTF}\right] \quad (2.68)$$

where $\Delta W_{vib}(T,\xi^{\ddagger})$ corrects $W(T,\xi^{\ddagger})$ to account for quantized vibrations orthogonal to the reaction coordinate (ξ) along the PMF at its maximum ($\xi^{\ddagger}$). $\Delta W_{vib,RS}(T)$ corrects $W(T,\xi_{RS})$ for quantized vibrations at the reactant minimum of the PMF (ξ_{RS}) and G_{RTF} is a correction for the vibrational free energy of the reactant mode that correlates with motion along the RC.[102] Vibrational corrections to the PMF can be calculated as the difference between the quantum an classical contributions at the TS and RS:

$$\begin{aligned}\Delta W_{vib} &= \Delta W_{vib}\left(T,\xi^{\ddagger}\right) - \Delta W_{vib,RS}\left(T\right) = \\ &= [ZPE_{TS} - ZPE_{RS}] + \\ &+ \left[\left\langle \sum_{i=1}^{3N-7} RTln\left(1 - e^{\frac{h\omega_i}{k_bT}}\right)\right\rangle_{TS} - \left\langle \sum_{i=1}^{3N-6} RTln\left(1 - e^{\frac{h\omega_i}{k_bT}}\right)\right\rangle_{RS}\right] - \\ &- \left[\left\langle \sum_{i=1}^{3N-7} RTln\frac{h\omega_i}{k_bT}\right\rangle_{TS} - \left\langle \sum_{i=1}^{3N-6} RTln\frac{h\omega_i}{k_bT}\right\rangle_{RS}\right]\end{aligned} \quad (2.69)$$

where N_A is the Avogadro number and c is the speed of light. The term inside the first square bracket refers to the calculation of the ZPE difference between RS and TS, while the second and the last ones are the differences between quantum and classical thermal vibrational free energies, respectively. However, despite the advances in the quantum formulation of rate constants, the study reactions involving the transfer

of light atoms like proton transfer reaction presents many challenges, not only because of the quantum character of those particles that require the quantization of their movements to reproduce correctly their behavior, but also because of the coupling between the solvent molecules and vibrational modes inside the reactive complex with the transferred atom, since the donor-acceptor distance vibrational mode is a typically dominant feature in H-atom transfers.[105-107]

Borgis-Hynes Theory

In the "traditional" view of light particle transfer reactions, the RC includes explicitly classical motions of the transferred atom. Then a tunnelling correction is added to the rate expression to account for the transmission probability through the reaction barrier, as seen in equation (2.69).[108-113] On the other hand, in the "non-traditional" picture,[65,74,76-79,114-126] the reaction is driven by configurational changes in the surrounding environment and the activation free energy is largely determined by the environmental reorganization. Different states of solvation distort the instantaneous potential, felt by the transferred atom along the transfer coordinate.

Borgis and Hynes[74,76,77,79] developed an analytic approach for the adiabatic and non-adiabatic proton transfer reaction rate constant, although it can be applied to similar reactions such as hydride transfer. The non-adiabatic limit refers to proton transfer in weakly interacting inter or intramolecular complexes for which the reaction process will be dominated by the tunneling of the proton while, in the adiabatic limit, the transfer of the light particle takes part without barrier and therefore without tunneling (see Figure 2.6b). This theory is based on a dynamic

and coupled description of the light particle motion and the vibrational modes of the substrate and the environment,[74] that provides an accurate route for including tunneling in the rate constant calculations.[76,79]

Couched in a Landau-Zener curve crossing approach[127] and inspired by the Marcus theory for electronic transfer reactions,[128-131] this approach relies on the Born-Oppenheimer approximation,[74,75] according to which time scale for the light particle motion is short compared to those on the environmental motions and other intramolecular vibrations. While the reactive system is described by the proton coordinate ($q_H = \big(d(D,H) - d(A,H)\big)$) and the Donor-Acceptor distance ($Q = d(D,A)$), the rest of degrees of freedom of the solute and the solvent are included in a general Solvent Coordinate (S). Hence, for a fixed Q and S coordinates, the motion of the proton is described by an electronically adiabatic double-well potential $V(q_H, Q, S)$. The reaction evolves with the atoms in the S coordinate and moves from a configuration that favors the location of the proton close to the donor atom to a configuration where the proton is more probably located bonded to the acceptor. From one configuration to the other, the system passes through an intermediate point where the probability is the same for both cases (see Figure 2.6a).

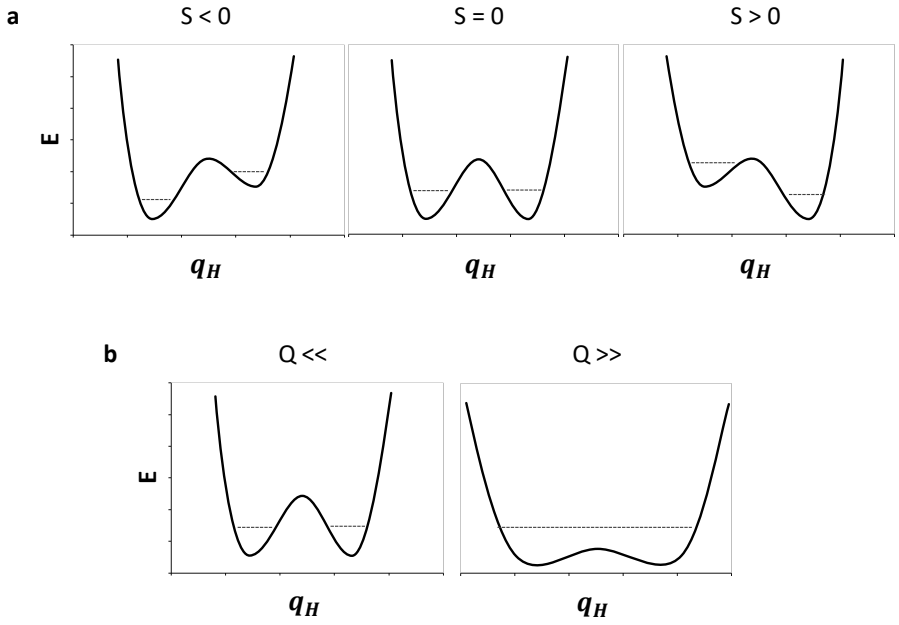


Figure 2.6: *Schematic proton transfer potentials vs. proton coordinate q_H for different values of the S coordinate (a) and of the Q coordinate (b). The diabatic proton vibrational levels are indicated.*

If we consider the energy of the H atom bonded to the donor and its energy when it is bonded to the acceptor, the reaction can be considered as a curve crossing problem. In this description, the proton transfer is a non-adiabatic event between the localized reactant (R) and product (P) states (see Figure 2.7).

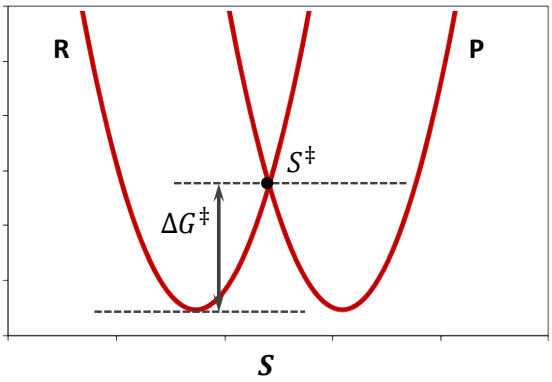


Figure 2.7: *Diabatic free energy curves in the solvent coordinate (S) for the reactants (R) and products (P) vibrational ground states in a symmetric reaction.*

The rate constant can be expressed as the average one-way flux (from R to P) in the solvent coordinate through the crossing point $S^{\ddagger}$ of the two free energy curves, with the inclusion of a general Landau-Zener transmission coefficient[127] (κ_{LZ}) giving the probability of a successful curve crossing: [79]

$$k_r = \langle \dot{S} \cdot \theta(\dot{S}) \cdot \delta(S - S^{\ddagger}) \cdot \kappa_{LZ}(\dot{S}) \rangle_R \tag{2.70}$$

$$\kappa_{LZ}(\dot{S}) = \frac{1 - e^{-\frac{2\pi \cdot C^2}{\hbar \cdot \dot{S}}}}{1 - \frac{1}{2} e^{-\frac{2\pi \cdot C^2}{\hbar \cdot \dot{S}}}} \tag{2.71}$$

where $\theta(\dot{S})$ is the velocity step function, whose thermal average is $\langle \theta(\dot{S}) \rangle = {}^1/_2$,[79] and C is the coupling between the proton ground state in the R and P wells. The coupling is exponentially sensitive to the Q coordinate, increasing as Q decreases.[79,126]

$$C(Q) = C_0 exp[-\alpha(Q - Q_0)] \tag{2.72}$$

being α a decay parameter, Q_0 and C_0 an equilibrium position and the coupling at that position respectively. The Q coordinate also controls the character of the barrier and how the proton is transferred. Namely for large equilibrium values of Q, the barrier in the q_H will become large enough so that the atom has to tunnel through it, while at smaller equilibrium values of Q, the barrier is reduced so that non-tunneling adiabatic transfer occurs.

The average is over the classical solvent distribution normalized by the partition distribution function of the solvent in the reactant region:

$$k_r = g(C) \cdot e^{-\frac{\Delta G^{\ddagger}_{ZPE}}{RT}} \tag{2.73}$$

where $g(C)$ is:

$$g(C) = \int \left(d\dot{S} \cdot \kappa_{LZ}(C) \cdot \dot{S} \cdot \theta(\dot{S}) \cdot e^{-\frac{m_S \dot{S}^2}{2 \cdot RT}} \right) \cdot \frac{\sqrt{k_s \cdot m_s}}{2\pi \cdot RT} \tag{2.74}$$

and k_S is the force constant of the solvent coordinate, which is calculated from its frequency (ω_S) and effective mass (m_S): $k_S = m_S \cdot \omega_S^2$.

In this case $\Delta G^{\ddagger}_{ZPE}$ is the free energy of activation with a correction to the barrier of the Zero Point Energy (ZPE); i.e. the vibrational ground state; of the light particle in the reactant and product states.

$$\Delta G^{\ddagger}_{ZPE} = \Delta G^{\ddagger} + \Delta ZPE \tag{2.75}$$

$$\Delta ZPE = ZPE_{TS} - ZPE_{RS} \tag{2.76}$$

The ZPE can be calculated with the harmonic approximation:[132]

$$\Delta ZPE = \left\langle \sum_{i=1}^{3N-7} \frac{N_A \cdot c \cdot \hbar\omega_i}{2} \right\rangle_{TS} - \left\langle \sum_{i=1}^{3N-6} \frac{N_A \cdot c \cdot \hbar\omega_i}{2} \right\rangle_{RS} \tag{2.77}$$

where N_A is the Avogadro number, c is the speed of light and ω_i the normal modes of the system in the RS and TS.

As we mentioned previously, the reaction can happen in two different limit regimes, in function of the degree of coupling between the two states. If it is very low, we are at the non-adiabatic limit and κ_{LZ} can be approximated to:

$$\kappa_{LZ} \approx \frac{4\pi C^2}{\hbar \dot{S}} \tag{2.78}$$

Thus the rate constant will be:

$$k_r = \left(\frac{2\pi k_S}{RT}\right)^{\frac{1}{2}} \cdot \frac{C^2}{\hbar} \cdot e^{-\frac{\Delta G^{\ddagger}_{ZPE}}{RT}} \tag{2.79}$$

On the other hand, if the coupling is too high, κ_{LZ} will be close to unity, we will be at the adiabatic level and the rate constant can be calculated as:

$$k_r = \frac{\omega_s}{2\pi} \cdot e^{-\frac{\Delta G^{\ddagger}}{RT}} \tag{2.80}$$

which is equivalent to equation (2.59) considering that the activation free energy is defined for TS and RS placed at the maximum and minimum of the free energy profile and there are ensembles with the same degrees of freedom.[79]

Quality of the Transition State Ensemble

Classical theories based on TST for reactions in solution usually try to optimize the dividing surface and then reduce substantially the solvent-induced recrossing that can produce those degrees of freedom that deviate from equilibrium. The committor is a test performed to control the optimization of the TS ensemble carrying molecular dynamic trajectories starting from structures that belong to the TS ensemble with different random initial velocities taken from a Maxwell–Boltzmann distribution. If the TS ensemble is ideal, the probability for the system of ending in reactants of products will be the same. So, the quality of the ensemble will be determined by the number of structures that have a 0.5 of probability of ending in products. If we use the averaged probability of every structure to end in products, a result close to 0.5 usually means that we have a good definition of the ensemble. Nevertheless, if half the number of structures got a probability of 1 to end in products and the other half got a 0 probability, the total average would also be 0.5, but the ensemble definition would be completely wrong. That is the reason why the committor is usually given as whole probability histogram (see Figure 2.8).

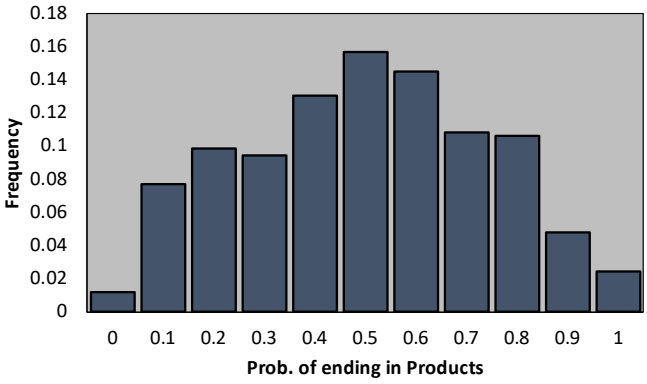


Figure 2.8: *Example of a committor histogram used to check the quality of the TS ensemble. In this case, as most of the structures are distributed around the 0.5 probability value, we would have an adequate definition of the TS ensemble.*

Transmission Coefficient

Another way to characterize the TS ensemble is by calculating the transmission coefficient. This can be done in a similar way to the committor analysis. Trajectories form the TS are now performed for positive ($t > 0$) and negative ($t < 0$) times reversing the initial velocities. If N trajectories are obtained from an equilibrium distribution we can consider that all of them have the same probability. Then, from these trajectories we can compute the transmission coefficient as:[63]

$$\kappa = \frac{\sum_{i=1}^{N} \boldsymbol{v_i} Q_i}{\sum_{i=1}^{N} |\boldsymbol{v_i}|} \tag{2.81}$$

where v_i is the initial velocity associated with the reaction coordinate, and Q_i is equal to 1 for reactant to product trajectories, 0 for reactant to reactant or product to product, and −1 for product to reactant

trajectories. The closer the value to unity the better the quality of the TS ensemble.

Bibliography

(1) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.

(2) Ponder, J. W.; Richards, F. M. *J. Comput. Chem.* **1987**, *8*, 1016.

(3) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.

(4) Cornell, W. D.*et al. J. Am. Chem. Soc.* **1995**, *117*, 5179.

(5) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.

(6) Ames, J. B.; Ishima, R.; Tanaka, T.; Gordon, J. I.; Stryer, L.; Ikura, M. *Nature* **1997**, *389*, 198.

(7) Kollman, P. A.*et al. Acc. Chem. Res.* **2000**, *33*, 889.

(8) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474.

(9) Patel, S.; Mackerell, A. D.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1504.

(10) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656.

(11) Allen, M. P.; Tildesley, D. J. *Computer simulation of liquids*; Oxford university press: New York, **1987**.

(12) Petraglio, G.; Ceccarelli, M.; Parrinello, M. *J. Chem. Phys.* **2005**, *123*, 044103.

(13) Field, M. J. *A Practical Introduction to the Simulation of Molecular Systems*; Cambridge University Press, **1999**.

(14) Ewald, P. P. *Ann. Phys.* **1921**, *369*, 253.

(15) Kuwajima, S.; Warshel, A. *J. Chem. Phys.* **1988**, *89*, 3751.

(16) Born, M.; Oppenheimer, R. *Annalen der Physik* **1927**, *389*, 457.

(17) Mills, I. M. *J. Phys. Chem.* **1984**, *88*, 532.

(18) Fukui, K.; Kato, S.; Fujimoto, H. *J. Am. Chem. Soc.* **1975**, *97*, 1.

(19) Fukui, K. *Acc. Chem. Res.* **1981**, *14*, 363.

(20) Warshel, A.; Karplus, M. *J. Am. Chem. Soc.* **1972**, *94*, 5612.

(21) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.

(22) Levitt, M. *J. Mol. Biol.* **1976**, *104*, 59.

(23) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718.

(24) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700.

(25) Théry, V.; Rinaldi, D.; Rivail, J. L.; Maigret, B.; Ferenczy, G. G. *J. Comput. Chem.* **1994**, *15*, 269.

(26) Monard, G.; Loos, M.; Théry, V.; Baka, K.; Rivail, J. L. *Int. J. Quant. Chem.* **1996**, *58*, 153.

(27) Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem. A* **2000**, *104*, 1720.

(28) Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467.
(29) Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714.
(30) Karplus, M.; McCammon, J. A. *Nat. Struct. Mol. Biol.* **2002**, *9*, 646.
(31) Kanaan, N.; Ruiz-Pernía, J. J.; Williams, I. H. *Chem. Commun.* **2008**, 6114.
(32) Warshel, A. *Proc. Nat. Acad. Sci. USA* **1984**, *81*, 444.
(33) Neria, E.; Karplus, M. *Chem. Phys. Lett.* **1997**, *267*, 23.
(34) García-Viloca, M.; Poulsen, T. D.; Truhlar, D. G.; Gao, J. *Protein Sci.* **2004**, *13*, 2341.
(35) Roca, M.; Andrés, J.; Moliner, V.; Tuñón, I.; Bertrán, J. *J. Am. Chem. Soc.* **2005**, *127*, 10648.
(36) Roca, M.; Moliner, V.; Tuñón, I.; Hynes, J. T. *J. Am. Chem. Soc.* **2006**, *128*, 6186.
(37) Ruiz-Pernía, J. J.; Tuñón, I.; Moliner, V.; Hynes, J. T.; Roca, M. *J. Am. Chem. Soc.* **2008**, *130*, 7477.
(38) Roca, M.; Oliva, M.; Castillo, R.; Moliner, V.; Tuñón, I. *Chem. Eur. J.* **2010**, *16*, 11399.
(39) Ruiz-Pernía, J. J.; Luk, L. Y. P.; García-Meseguer, R.; Martí, S.; Loveridge, E. J.; Tuñón, I.; Moliner, V.; Allemann, R. K. *J. Am. Chem. Soc.* **2013**, *135*, 18689.
(40) García-Meseguer, R.; Martí, S.; Ruiz-Pernía, J. J.; Moliner, V.; Tuñón, I. *Nat. Chem.* **2013**, *5*, 566.
(41) García-Meseguer, R.; Zinovjev, K.; Roca, M.; Ruiz-Pernía, J. J.; Tuñón, I. *J. Phys. Chem. B* **2015**, *119*, 873.
(42) Verlet, L. *Phys. Rev.* **1967**, *159*, 98.
(43) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. *J. Chem. Phys.* **1982**, *76*, 637.
(44) Brünger, A.; Brooks III, C. L.; Karplus, M. *Chem. Phys. Lett.* **1984**, *105*, 495.
(45) Wu, X.; Brooks, B. R. *Chem. Phys. Lett.* **2003**, *381*, 512.
(46) McQuarrie, D. A. *Statistical Mechanics, Harper &*; Harper & Row: New York, **1976**.
(47) Beveridge, D. L.; DiCapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431.
(48) Jorgensen, W. L. *Acc. Chem. Res.* **1989**, *22*, 184.
(49) Straatsma, T. P.; McCammon, J. A. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407.
(50) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395.
(51) Kästner, J.; Senn, H. M.; Thiel, S.; Otte, N.; Thiel, W. *J. Chem. Theory Comput.* **2006**, *2*, 452.
(52) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300.

(53) Roux, B. **1995**, *91*, 275.

(54) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187.

(55) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011.

(56) Frenkel, D.; Smit, B. *Understanding Molecular Simulation*; Academic Press: San Diego, **2002**.

(57) Arrhenius, S. *Z. Phys. Chem.* **1889**, *4*, 226.

(58) Pelzer, H.; Wigner, E. *Zeitschrift für Physikalische Chemie, Abteilung B: Chemie der Elementarprozesse, Aufbau der Materie* **1932**, *15*, 445.

(59) Eyring, H. *J. Chem. Phys.* **1935**, *3*, 107.

(60) Evans, M. G.; Polanyi, M. *Trans. Far. Soc.* **1935**, *31*, 875.

(61) Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J. *J. Phys. Chem.* **1996**, *100*, 12771.

(62) Wigner, E. *Trans. Far. Soc.* **1938**, *34*, 29.

(63) Bergsma, J. P.; Gertner, B. J.; Wilson, K. R.; Hynes, J. T. *J. Chem. Phys.* **1987**, *86*, 1356.

(64) Gertner, B. J.; Bergsma, J. P.; Wilson, K. R.; Lee, S.; Hynes, J. T. *J. Chem. Phys.* **1987**, *86*, 1377.

(65) Gertner, B. J.; Wilson, K. R.; Zichi, D. A.; Lee, S.; Hynes, J. T. *Faraday Discuss. Chem. Soc.* **1988**, *85*, 297.

(66) Gertner, B. J.; Wilson, K. R.; Hynes, J. T. *J. Chem. Phys.* **1989**, *90*, 3537.

(67) Martí, S.; Roca, M.; Andres, J.; Moliner, V.; Silla, E.; Tunon, I.; Bertran, J. *Chem. Soc. Rev.* **2004**, *33*, 98.

(68) Soriano, A.; Silla, E.; Tuñón, I.; Martí, S., S.; Moliner, V.; Bertrán, J. *Theor. Chem. Accounts* **2004**, *112*, 327.

(69) Soriano, A.; Silla, E.; Tuñón, I.; Ruiz-López, M. F. *J. Am. Chem. Soc.* **2005**, *127*, 1946.

(70) Roca, M.; Moliner, V.; Ruiz-Pernía, J. J.; Silla, E.; Tuñón, I. *J. Phys. Chem. B* **2005**, *110*, 503.

(71) Schenter, G. K.; Garrett, B. C.; Truhlar, D. G. *J. Chem. Phys.* **2003**, *119*, 5828.

(72) Mullen, R. G.; Shea, J. E.; Peters, B. *J. Chem. Phys.* **2014**, *140*, 041104.

(73) Mullen, R. G.; Shea, J. E.; Peters, B. *J. Chem. Theory Comput.* **2014**, *10*, 659.

(74) Borgis, D. C.; Lee, S.; Hynes, J. T. *Chem. Phys. Lett.* **1989**, *162*, 19.

(75) Borgis, D.; Hynes, J. T. In *The Enzyme Catalysis Process: Energetics, Mechanism and Dynamics*; Cooper, A., Houben, J., Chien, L., Eds.; Springer US: New York, **1989**, p 293.

(76) Borgis, D. C.; Hynes, J. T. *J. Chem. Phys.* **1991**, *94*, 3619.

(77) Borgis, D.; Hynes, J. T. *Chem. Phys.* **1993**, *170*, 315.

(78) Staib, A.; Borgis, D.; Hynes, J. T. *J. Chem. Phys.* **1995**, *102*, 2487.

(79) Borgis, D. C.; Hynes, J. T. *J. Phys. Chem.* **1996**, *100*, 1118.

(80) Glasstone, S.; Laidler, K. J.; Eyring, H. *The theory of rate processes: the kinetics of chemical reactions, viscosity, diffusion and electrochemical phenomena*; McGraw-Hill Book Company, inc.: New York, **1941**.

(81) Eyring, H.; Polanyi, M. *Z Phys Chem Abt B* **1931**, *12*, 279.

(82) Eyring, H.; Polanyi, M. In *Zeitschrift für Physikalische Chemie* **2013**; Vol. 227, p 1221.

(83) Pollak, E. In *Activated Barrier Crossing: applications in physics, chemistry and biology*; Fleming, G. R., Häggi, P., Eds.; World Scientific Publishyng Co. Pte. Ltd.l: Singapore, **1993**, p 5.

(84) Truhlar, D. G.; Garrett, B. C. *Acc. Chem. Res.* **1980**, *13*, 440.

(85) Tuñón, I.; Laage, D.; Hynes, J. T. *Arch. Biochem. Biophys.* **2015**, *582*, 42.

(86) Grote, R. F.; Hynes, J. T. *J. Chem. Phys.* **1980**, *73*, 2715.

(87) Ford, G. W.; Kac, M.; Mazur, P. *J. Math. Phys.* **1965**, *6*, 504.

(88) Adelman, S. A.; Doll, J. D. *J. Chem. Phys.* **1976**, *64*, 2375.

(89) Cortés, E.; West, B. J.; Lindenberg, K. *J. Chem. Phys.* **1985**, *82*, 2708.

(90) Messiah, A. In *Quantum Mechanics*; North-Holland: Amsterdam, **1958**; Vol. 1, p 299.

(91) Miller, W. H. *J. Phys. Chem.* **1974**, *61*, 1823.

(92) García-Viloca, M.; Alhambra, C.; Truhlar, D. G.; Gao, J. *J. Chem. Phys.* **2001**, *114*, 9953.

(93) Meana-Pañeda, R.; Fernández-Ramos, A. In *Kinetics and Dynamics: From Nano- to Bio-Scale*; Paneth, P., Dybala-Defratyka, A., Eds.; Springer Netherlands: Dordrecht, **2010**, p 481.

(94) Kuppermann, A.; Truhlar, D. G. *J. Am. Chem. Soc.* **1971**, *93*, 1840.

(95) Marcus, R. A. *J. Chem. Phys.* **1966**, *45*, 4493.

(96) Marcus, R. A.; Coltrin, M. E. *J. Chem. Phys.* **1977**, *67*, 2609.

(97) Skodje, R. T.; Truhlar, D. G.; Garrett, B. C. *J. Chem. Phys.* **1982**, *77*, 5955.

(98) Lu, D.*et al. Comput. Phys. Commun.* **1992**, *71*, 235.

(99) Garrett, B. C.; Truhlar, D. G.; Wagner, A. F.; Dunning, T. H. *J. Chem. Phys.* **1983**, *78*, 4400.

(100) Garrett, B. C.; Abusalbi, N.; Kouri, D. J.; Truhlar, D. G. *J. Chem. Phys.* **1985**, *83*, 2252.

(101) Fernandez-Ramos, A.; Truhlar, D. G. *J. Chem. Phys.* **2001**, *114*, 1491.

(102) Alhambra, C.; Corchado, J.; Sánchez, M. L.; García-Viloca, M.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. B* **2001**, *105*, 11326.

(103) Truhlar, D. G.; Gao, J.; Alhambra, C.; García-Viloca, M.; Corchado, J.; Sánchez, M. L.; Villà, J. *Acc. Chem. Res.* **2002**, *35*, 341.

(104) Truhlar, D. G.; Gao, J.; García-Viloca, M.; Alhambra, C.; Corchado, J.; Luz Sánchez, M.; Poulsen, T. D. *Int. J. Quant. Chem.* **2004**, *100*, 1136.

(105) Trakhtenberg, L. I.; Klochikhin, V. L.; Pshezhetsky, S. Y. *Chem. Phys.* **1982**, *69*, 121.

(106) Siebrand, W.; Wildman, T. A.; Zgierski, M. Z. *J. Am. Chem. Soc.* **1984**, *106*, 4083.

(107) Siebrand, W.; Wildman, T. A.; Zgierski, M. Z. *J. Am. Chem. Soc.* **1984**, *106*, 4089.

(108) Westheimer, F. H. *Chem. Rev.* **1961**, *61*, 265.

(109) Hwang, J. K.; Warshel, A. *J. Phys. Chem.* **1993**, *97*, 10053.

(110) Hwang, J. K.; Warshel, A. *J. Am. Chem Soc.* **1996**, *118*, 11745.

(111) Alhambra, C.; Gao, J.; Corchado, J. C.; Villà, J.; Truhlar, D. G. *J. Am. Chem Soc.* **1999**, *121*, 2253.

(112) Alhambra, C.; Corchado, J. C.; Sánchez, M. L.; Gao, J.; Truhlar, D. G. *J. Am. Chem. Soc* **2000**, *122*, 8197.

(113) Cui, Q.; Karplus, M. *J. Am. Chem. Soc.* **2002**, *124*, 3093.

(114) Timoneda, J. J.; Hynes, J. T. *J. Phys. Chem.* **1991**, *95*, 10431.

(115) Andot, K.; Hynes, J. T. *J. Mol. Liq.* **1995**, *64*, 25.

(116) Ando, K.; Hynes, J. T. *Faraday Discuss.* **1995**, *102*, 435.

(117) Ando, K.; Hynes, J. T. *J. Phys. Chem. B* **1997**, *101*, 10464.

(118) Ando, K.; Hynes, J. T. *J. Phys. Chem. A* **1999**, *103*, 10398.

(119) Kiefer, P. M.; Hynes, J. T. *J. Phys. Chem. A* **2002**, *106*, 1834.

(120) Kiefer, P. M.; Hynes, J. T. *J. Phys. Chem. A* **2002**, *106*, 1850.

(121) Kiefer, P. M.; Hynes, J. T. *J. Phys. Chem. A* **2003**, *107*, 9022.

(122) Kiefer, P. M.; Hynes, J. T. *J. Phys. Chem. A* **2004**, *108*, 11809.

(123) Kiefer, P. M.; Hynes, J. T. *J. Phys. Chem. A* **2004**, *108*, 11793.

(124) Ando, K.; Hynes, J. T. In *Advances in Chemical Physics*; John Wiley & Sons, Inc.: **2007**, p 381.

(125) Kiefer, P. M.; Hynes, J. T. In *Hydrogen-Transfer Reactions*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim **2007**, p 303.

(126) Kiefer, P. M.; Hynes, J. T. *J. Phys. Org. Chem.* **2010**, *23*, 632.

(127) Newton, M. D.; Sutin, N. *Annu. ReV. Phys. Chem.* **1984**, *35*, 437.

(128) Marcus, R. A. *J. Chem. Phys.* **1956**, *24*, 966.

(129) Marcus, R. A. *J. Chem. Phys.* **1957**, *26*, 867.

(130) Marcus, R. A. *J. Chem. Phys.* **1957**, *26*, 872.

(131) Marcus, R. A. *Annu. Rev. Phys. Chem.* **1964**, *15*, 155.

(132) Griffiths, D. J. In *Introduction to quantum mechanics*; Challice, J., Fahlgren, E., Botting, C., Eds.; Pearson Prentice Hall: Upper Saddle River, **1995**.

Chapter 3
Environmental Coordinates

In the context of the present work we define the environment as the ensemble of atoms that do not participate directly in the reaction, whereas the solute is defined as the atoms or molecules directly involved in the reaction process. Thereby, we can define environmental coordinates as those reaction coordinates that include in their definitions most of the degrees of freedom of the environment.

The main objective of the present project has been the development and implementation of two environmental coordinates that can solve some of the problems that appear with the application of TST to the study of chemical reactions. This chapter will include a brief introduction to the problems, which have been only alluded to in the previous chapter, and will be dealt with in this one. Then, we will present a brief explanation of environmental coordinates and finally we will provide a theoretical description of the ones developed and employed in this thesis project.

3.1. The Protein Dynamics Problem

It is well known that, in order to function, enzymes need to be flexible enough to permit the evolution among the different conformations relevant at each step of the full catalytic process.[1,2] Even within an active site designed to accommodate the charge distribution of the TS,[3] some motions in protein coordinates are needed to evolve from reactant configurations to the TS ensemble.[2,4-6] However, the impact of protein dynamics on the rate constant of the chemical step remains the subject of a long-standing debate in scientific literature.[5-18] The participation of protein motions in the progress of the reaction is a well-established fact.[6,7,10,19] Still, an important question has to be address: can the impact of protein motions on the rate constant be described with the current theoretical frameworks –based in TST– to explain adequately the rate of enzymatic reactions?[15,19]

Certainly there are dynamic effects coupled with the chemical reactions at a molecular level. For example, a characteristic feature of enzymatic environments is the presence of a very broad spectrum of protein conformational motions, which occur on timescales ranging from picoseconds to milliseconds.[20] But are these dynamics relevant for the reaction rates?

Large amplitude conformational changes in the protein may occur during substrate binding and product release, which is not directly involved in the chemical process. Nevertheless, our focus in this thesis is in the chemical transformations themselves. Here, it is expected that faster conformational changes occur in the active site with similar timescales of

the chemical transformation. These changes affect the interaction between the protein active site and the substrate. One possible effect would be to change the electrostatic properties and the hydrogen-bond network of the active site in order to favour the electronic rearrangement associated with the bond-breaking and bond-forming processes.

As mentioned previously, TST provides the tools for calculating the rate of the reaction, without any explicit accounting for dynamical effects. However, it is easy to imagine they might play a role here. The reacting fragments must be moved close together and correctly oriented; some bonds must be lengthened while others are shortened.[21] As well, conformational changes are needed to tune the electrostatics to favour the process.[22,23] Indeed, dynamics can also be found within the TS region itself. For example, large amplitude motions might be opposed by dynamic viscous environment forces. Charge transfers and redistributions might be opposed by sluggish motional rearrangement of polar and charged molecules in the environment.[24,25]

A popular and widespread criterion for the presence of dynamical effects is the departure of the reaction rate constant from its TST value,[24] conveniently measured by the transmission coefficient (κ) defined by the ratio k/k_{TST}, being k the rate constant of the reaction and k_{TST} the rate constant calculated using TST. In this view, the more κ is reduced, the greater the dynamical effects on the rate constant are. However, κ depends on the definition of the RC in the TS neighbourhood.[25] Thus the transmission coefficient value could be unity (or very close to it) for one choice of RC, but very small for another choice, with a corresponding

difference in the assessment of the importance of dynamics. Another problematic issue that may arise occurs when κ is lower than unity, but no environmental dynamics is involved, in fact, the absence of dynamics is responsible for the depression of κ.[26] In any condensed system it is assumed that the environment is always equilibrated to the instantaneous charge distribution in the reacting solute. This is an ''equilibrium solvation'' assumption and any deviation from it will produce the departure of κ from unity.

All the above discussion has assumed classical motion for the nuclei; in particular, the reaction coordinate for the passage through the TS surface is described by classical nuclear motions. However enzymatic reactions involving quantum tunnelling have provided an important area for discussion of possible dynamical effects. Tunnelling clearly depends on the barrier along the transfer coordinate, which, in turn, is extremely sensitive to the distance between the H-donor and H-acceptor, often called the 'promoting' or 'gating' mode, and this parameter can be coupled to various features of the enzyme. However, this dependence and coupling do not produce any dynamical effects on the reaction rate, provided that the donor–acceptor distance has an equilibrium distribution, but may change the rate of tunnelling or even change the quantum regime from tunneling to the quantum adiabatic regime.[24] Besides, as we mentioned previously in Chapter 2, these light particles are driven by the different configurations of the environment. Thus, an extensive environmental reorganization is involved in the reaction and is the source of an activation free energy for the rate constant.

For a better understanding of dynamical effects, we need to control the evolution of the environment and its behaviour along the chemical transformation. In these terms we propose two different solutions:

- One option is to project the multidimensional Free-Energy Surface (FES) of the enzymatic reaction in a 2D model obtained as a function of a solute coordinate and an environmental coordinate. Such FES will allow us to estimate the timing and coupling between the solute and the solvent motions along the reaction. Moreover, including the degrees of freedom of the solvent in the definition of the RC will reduce significantly the effects of the nonequilibrium solvation over the rate constant, hence reducing recrossings and minimizing the impact on the transmission coefficient.

- The second solution is to define the RC in terms of the environmental reorganization, assuming that light atoms involved in the reaction adapt to the environment. This solution will also allow us to include quantum properties of the light particles in the estimation of the rate constant.

3.2. Energy Gap Coordinate

An RC definition that takes into account the environmental degrees of freedom is given by the empirical valence bond (EVB) formalism, developed by Warshel and colleagues.[27] Within this approach, the reaction is represented by a series of resonance states. Each resonance state is characterized by an empirical energy function, and the ground state energy of the system is obtained by solving the secular equation of the system similarly as in the valence bond theory.[28]

The reaction coordinate is defined, in the framework of the EVB method,[29] as a difference between the potential energies of the resonance states. This Energy Gap coordinate (E_{GAP}) has the advantage of being capable of describing the reorganization of the system toward the transition state, while it requires only the definition of the resonance states and not of the reaction pathway itself. This complex RC is able to capture all essential aspects of enzymatic reactions including the conformational changes that influence the energetics of the reaction.

In the EVB method,[27] energies of the resonance states (ε_{ii}) are approximated by a classical potential, constructed by the Morse terms for the breaking and forming bonds, the harmonic terms for bonds, angles, torsions for covalently bonded atoms and the non-bonded terms including electrostatic and van der Waals energies.[30]

For simplicity, only two resonance states will be considered here, although more can also be handled. The first resonance state is the

reactant, and the second state is the product of the reaction. Then, the Hamiltonian of the system is the following:

$$\boldsymbol{H} = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \tag{3.1}$$

where $H_{11} = \varepsilon_{11}$ and $H_{22} = \varepsilon_{22} + \alpha_2$, being ε_{11} and ε_{22} the energies of the first and second resonance states (reactants and products respectively) and $\alpha_2$the energy offset between two resonance states. The off-diagonal terms H_{12} and H_{21} are the interaction energies between resonance states that can be represented by a simple exponential function:

$$H_{12} = A_{12} \cdot e^{-\mu_{12} \cdot r_{ab}} \tag{3.2}$$

where r_{ab} is the representative distance between the atoms whose bonding is changed from first to second resonance structure, whereas A_{12} and μ_{12} are parameters.

In the framework of the valence bond method, a reaction can be considered as a transition from the reactant to product state. This transition can be characterized by a reaction coordinate E_{GAP} that is defined by the difference (gap) between the energies of the resonance states:

$$E_{GAP}(\boldsymbol{x}) = \varepsilon_{11}(\boldsymbol{x}) - \varepsilon_{22}(\boldsymbol{x}) \tag{3.3}$$

As E_{GAP} depends on all system coordinates ($\boldsymbol{x}$), it can in principle describe the reorganization of the whole system toward a transition state including the reorganization of the solvent and of the protein environment in enzymatic reactions. This feature is missing in reaction coordinates that are based on local geometrical parameters.

3.3. Electrostatic Coordinate

The E_{GAP} coordinate, where both solute and environment motions are incorporated, is based on the EVB formalism. However, in order to address the problem of the participation of environmental motions in QM/MM simulations a different approach is needed.

Studies in aqueous solution have already established that the reaction proceeds with an important change in the electrostatic potential created by the environment.[23,31] Besides, any fluctuations of the environment will change the electrostatic interactions with the reactive system.[23] Thus the electrostatic potential perceived by a given atom should be an adequate property to control the evolution of the environment along the reaction and can be used altogether as an RC to improve the definition of the TS ensemble.

This Electrostatic Coordinate (Ep) takes the electrostatic potential of the MM part, in a hybrid QM/MM system created on an atom directly involved in the reaction process, usually one that suffers significant charge redistribution along the chemical reaction, and computes it as a Solvent Coordinate.

$$Ep = \sum_{j=1}^{N} \frac{q_j}{|\boldsymbol{r_j} - \boldsymbol{r_A}|} \tag{3.4}$$

where A makes reference to the selected atom of the reacting system, q_j is the charge of the j atom, $\boldsymbol{r_j}$ and $\boldsymbol{r_A}$ the coordinates of both atoms.

The Ep coordinate can be defined as a combination of the electrostatic potential perceived by different atoms since, when one atom suffers substantial charge redistribution there is often another one that suffers the opposite redistribution. Thus, an antisymmetric combination of the electrostatic potential perceived by two atoms (A, B) or group of atoms will better represent the effect of the environment on the reaction.

$$Ep = V_A - V_B = \sum_{j=1}^{N} \frac{q_j}{|\boldsymbol{r_j} - \boldsymbol{r_A}|} - \sum_{j=1}^{N} \frac{q_j}{|\boldsymbol{r_j} - \boldsymbol{r_B}|} \tag{3.5}$$

The Ep controls the evolution of the environment along the reaction, so it can be used, in combination with a solute coordinate that controls the atoms directly involved in the reaction, as a proper environmental coordinate. Atoms that suffer charge redistribution along the chemical process are usually the most relevant for the reaction and thus, atoms A and B will also most probably be included in the solute coordinate. Therefore, the Ep coordinate will not be orthogonal to the solute coordinate, although the coupling between the two is expected to be small, as will be discussed in the next chapter.

3.4. Energy Difference Coordinate

Proton, hydrogen and hydride transfer are chemical reactions of undeniable importance in chemistry and biology.[32-34] The simulation of these reactions is challenging due to the importance of both nuclear and electronic quantum effects, as well as the effects that the environmental degrees of freedom can have in the reaction. The incorporation of quantum effects is required for the proper description of the transfer of a light particle.[35-46] Although the ideas presented below can be equally applied to proton, hydrogen and hydride transfer reactions, we will focus on the hydride transfer reaction, which is the subject of the present thesis. The energy difference coordinate ($\Delta\varepsilon$) presented here provides an excellent tool to account the quantum nuclear character of the motion of the light particles we want to quantize, using standard QM/MM packages, in contrast with the E_{GAP} coordinate, which is appropriate for EVB representations.

Based on the previous works done by Borgis and Hynes for proton transfer reactions,[36,37,47-49] this coordinate reflects the energetic interaction between the environmental atoms and the reacting solute system. The coordinates of atoms to be quantized are selected to define two reference states, whose energies are calculated at a frozen environmental configuration. Then, the difference between the total energy of the system in each defined state is used as an environmental coordinate.

Being $\boldsymbol{r}$ and $\boldsymbol{\zeta}$ the coordinates of the selected atoms and those for the rest of them respectively (see Figure 3.1); $\epsilon(\boldsymbol{\zeta}, \boldsymbol{r})$ is the total energy of the

system, $\Delta\varepsilon$ will be defined as the energy difference between two states determined by the coordinates of the selected atoms:

$$\Delta\varepsilon = \epsilon(\boldsymbol{\zeta}, \boldsymbol{r_1}) - \epsilon(\boldsymbol{\zeta}, \boldsymbol{r_2}) \tag{3.6}$$

where $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$ are the coordinates that the selected atoms should have in reactant and product states respectively.

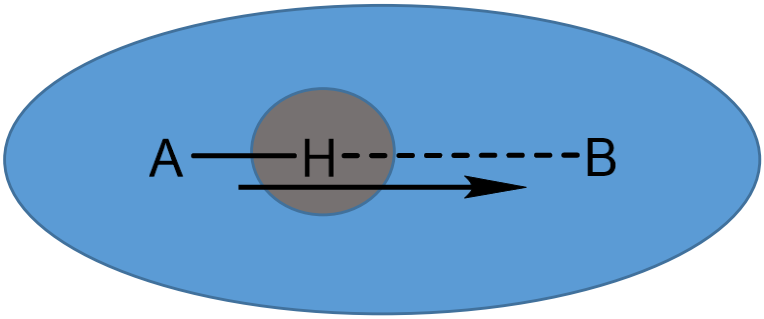


Figure 3.1: *Schematic representation of a hydrogen transfer reaction in condensed phase, the grey circle represents the coordinates of the selected atoms ($\boldsymbol{r}$), in this case one of hydrogen. The blue part depicts the coordinates of the rest of atoms ($\boldsymbol{\zeta}$) including A and B.*

The behaviour of the coordinate can be better understood if we follow the energy profile at different frozen environments along the transfer coordinate (q_H), i.e. the coordinate that defines the position of the selected atoms.

The behaviour of the coordinate ($\Delta\varepsilon$) along the reaction, as can be seen in Figure 3.2, will be:

a) At the RS $\boldsymbol{r_1}$ state has lower energy than $\boldsymbol{r_2}$ and thus $\Delta\varepsilon$ is negative.

b) At the TS both $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$ have the same energy and $\Delta\varepsilon$ is equal to zero.

c) Finally at the PS the opposite of the reactant case happens and $\Delta\varepsilon$ is positive.

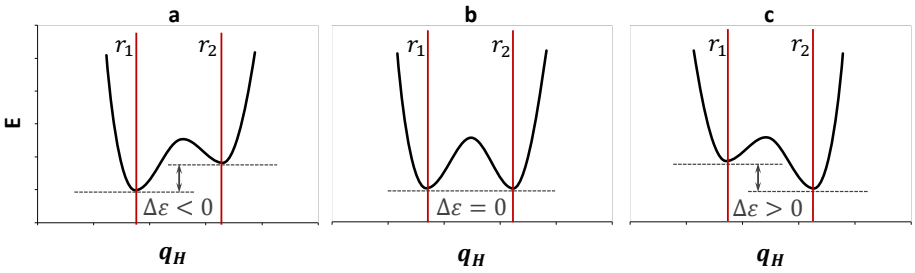


Figure 3.2: *Schematic representation of the instantaneous potential energy surface perceived by the selected atom along q_H at different coordinates of the environment. The red lines mark the point at where the states r_1 and r_2 are calculated. The grey dashed lines indicate the energy of each state.*

The $\Delta\varepsilon$ coordinate is orthogonal to the degrees of freedom of the selected atoms. So, assuming that its nuclear motions can be quantized at fixed values of the remaining slower environmental degrees of freedom;[36,37,39,44,45] the $\Delta\varepsilon$ coordinate will allow us to account the quantum nuclear character of the selected atoms.

This RC can be easily implemented in standard QM/MM packages to perform MD simulations at different values of $\Delta\varepsilon$ and then, calculate the free energy profile of the reaction. From these calculations it is possible to take snapshots of the trajectories and calculate the potential energy profile of the hydrogen atom at a given environmental configuration along the transfer coordinate. These profiles can be later used to calculate

the wavefunction of the quantized atoms employing the DVR methodology.[50,51]

Bibliography

(1) Henzler-Wildman, K. A.; Lei, M.; Thai, V.; Kerns, S. J.; Karplus, M.; Kern, D. *Nature* **2007**, *450*, 913.

(2) Hammes, G. G.; Benkovic, S. J.; Hammes-Schiffer, S. *Biochemistry* **2011**, *50*, 10422.

(3) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. *Chem. Rev.* **2006**, *106*, 3210.

(4) Karplus, M.; McCammon, J. A. *Ann. Rev. Biochem.* **1983**, *52*, 263.

(5) García-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Science* **2004**, *303*, 186.

(6) Kamerlin, S. C. L.; Warshel, A. *Proteins* **2010**, *78*, 1339.

(7) Antoniou, D.; Basner, J.; Núñez, S.; Schwartz, S. D. *Chem. Rev.* **2006**, *106*, 3170.

(8) Gao, J.; Ma, S.; Major, D. T.; Nam, K.; Pu, J.; Truhlar, D. G. *Chem. Rev.* **2006**, *106*, 3188.

(9) Olsson, M. H. M.; Parson, W. W.; Warshel, A. *Chem. Rev.* **2006**, *106*, 1737.

(10) Nashine, V. C.; Hammes-Schiffer, S.; Benkovic, S. J. *Curr. Opin. Chem. Biol.* **2010**, *14*, 644.

(11) Adamczyk, A. J.; Cao, J.; Kamerlin, S. C. L.; Warshel, A. *Proc. Nat. Acad. Sci. USA* **2011**, *108*, 14115.

(12) Bhabha, G.; Lee, J.; Ekiert, D. C.; Gam, J.; Wilson, I. A.; Dyson, H. J.; Benkovic, S. J.; Wright, P. E. *Science* **2011**, *332*, 234.

(13) Ramanathan, A.; Agarwal, P. K. *PLOS Biol.* **2011**, *9*, e1001193.

(14) Zhang, J.; Klinman, J. P. *J. Am. Chem. Soc.* **2011**, *133*, 17134.

(15) Glowacki, D. R.; Harvey, J. N.; Mulholland, A. J. *Nat. Chem.* **2012**, *4*, 169.

(16) García-Meseguer, R.; Martí, S.; Ruiz-Pernía, J. J.; Moliner, V.; Tuñón, I. *Nat. Chem.* **2013**, *5*, 566.

(17) Ruiz-Pernía, J. J.; Luk, L. Y. P.; García-Meseguer, R.; Martí, S.; Loveridge, E. J.; Tuñón, I.; Moliner, V.; Allemann, R. K. *J. Am. Chem. Soc.* **2013**, *135*, 18689.

(18) García-Meseguer, R.; Zinovjev, K.; Roca, M.; Ruiz-Pernía, J. J.; Tuñón, I. *J. Phys. Chem. B* **2015**, *119*, 873.

(19) Hay, S.; Scrutton, N. S. *Nat. Chem.* **2012**, *4*, 161.

(20) Henzler-Wildman, K.; Kern, D. *Nature* **2007**, *450*, 964.

(21) Shurki, A.; Štrajbl, M.; Villà, J.; Warshel, A. *J. Am. Chem. Soc.* **2002**, *124*, 4097.

(22) Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W. *B. B. A.-Proteins Proteom.* **2006**, *1764*, 1647.

(23) Ruiz-Pernía, J. J.; Martí, S.; Moliner, V.; Tuñón, I. *J. Chem. Theory Comput.* **2012**, *8*, 1532.

(24) Tuñón, I.; Laage, D.; Hynes, J. T. *Arch. Biochem. Biophys.* **2015**, *582*, 42.

(25) Pu, J.; Gao, J.; Truhlar, D. G. *Chem. Rev.* **2006**, *106*, 3140.

(26) Bergsma, J. P.; Gertner, B. J.; Wilson, K. R.; Hynes, J. T. *J. Chem. Phys.* **1987**, *86*, 1356.

(27) Warshel, A.; Weiss, R. M. *J. Am. Chem. Soc.* **1980**, *102*, 6218.

(28) Heitler, W.; London, F. *Z. Phys.* **1927**, *44*, 455.

(29) Warshel, A. *Computer modeling of chemical reactions in enzymes and solutions*; Wiley: New York, **1991**.

(30) Mones, L.; Kulhánek, P.; Simon, I.; Laio, A.; Fuxreiter, M. *J. Phys. Chem. B* **2009**, *113*, 7867.

(31) Hwang, J. K.; King, G.; Creighton, S.; Warshel, A. *J. Am. Chem. Soc.* **1988**, *110*, 5297.

(32) Kresge, A. J. *Acc. Chem. Res.* **1975**, *8*, 354.

(33) Hibbert, F. *AdV. Phys. Org. Chem.* **1987**, *Volume 22*, 113.

(34) Hibbert, F.; Emsley, J. *AdV. Phys. Org. Chem.* **1991**, *Volume 26*, 255.

(35) Dogonadze, R. R.; Kuznetsov, A. M.; Zakaraya, M. G.; Ulstrup, J. In *Tunneling in Biological Systems*; Sutin, B. C. A. M. C. D. R. S. F., Ed.; Academic Press: **1979**, p 145.

(36) Borgis, D.; Hynes, J. T. *Chem. Phys.* **1993**, *170*, 315.

(37) Staib, A.; Borgis, D.; Hynes, J. T. *J. Chem. Phys.* **1995**, *102*, 2487.

(38) Bahnson, B. J.; Klinman, J. P. In *Meth. Enzymology*; Academic Press: **1995**; Vol. Volume 249, p 373.

(39) Ando, K.; Hynes, J. T. *J. Phys. Chem. B* **1997**, *101*, 10464.

(40) Kohen, A.; Klinman, J. P. *Acc. Chem. Res.* **1998**, *31*, 397.

(41) Hammes-Schiffer, S. *Chem. Phys. Chem.* **2002**, *3*, 33.

(42) Truhlar, D. G.; Gao, J.; Alhambra, C.; García-Viloca, M.; Corchado, J.; Sánchez, M. L.; Villà, J. *Acc. Chem. Res.* **2002**, *35*, 341.

(43) Kiefer, P. M.; Hynes, J. T. *J. Phys. Chem. A* **2004**, *108*, 11809.

(44) Kiefer, P. M.; Hynes, J. T. In *Hydrogen-Transfer Reactions*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim **2007**, p 303.

(45) Kiefer, P. M.; Hynes, J. T. *J. Phys. Org. Chem.* **2010**, *23*, 632.

(46) Truhlar, D. G. *J. Phys. Org. Chem.* **2010**, *23*, 660.

(47) Borgis, D. C.; Lee, S.; Hynes, J. T. *Chem. Phys. Lett.* **1989**, *162*, 19.

(48) Borgis, D. C.; Hynes, J. T. *J. Chem. Phys.* **1991**, *94*, 3619.

(49) Borgis, D. C.; Hynes, J. T. *J. Phys. Chem.* **1996**, *100*, 1118.

(50) Colbert, D. T.; Miller, W. H. *J. Chem. Phys.* **1992**, *96*, 1982.

(51) Thompson, W. H. *J. Chem. Phys.* **2003**, *118*, 1059.

Chapter 4
Results and Discussion

The methodologies explained and discussed before have been applied to four different enzymatic systems, whose results will be presented here.

Our objective with these procedures was to improve the understanding of enzymatic catalysis. Hence, we selected enzymatic systems that have been previously analyzed from experimental and theoretical perspectives, so we could compare these studies to our own results.

4.1. Haloalkane Dehalogenase

The xenobiotic compound 1,2-dicloroethane (DCE) has been, by far, the most synthetized chlorohydrocarbon, with an industrial production of more than 25 million ton per year all over the world.[1] Its main use is in the production of vinyl chloride monomer, a precursor to the synthesis of polyvinyl chloride. However DCE is carcinogen for humans and was

added to the "blacklist" of pollutants in some countries such as the USA and Canada.[2]

The bacteria *Xanthobacter autotrophicus* produces an enzyme, named haloalkane dehalogenase (DhlA), capable of hydrolyzing chlorine-carbon bonds from a wide range of halogenated compounds.[3,4] Many of these compounds are contaminants, which makes this enzyme a magnificent decontaminant in water treatments with this kind of pollutants.[3]

The three-dimensional structure of DhlA has been solved with a 1.15 Å resolution using X-ray crystallography and can be accessed from the Protein Data Bank.[5] The protein consists of two domains. The main one provides a scaffold for the active site, which is located in a predominantly hydrophobic cavity of 75 $Å^3$.[6] The only charged residues located in the cavity (Asp124, His289 and Asp260) form the catalytic triad presented in many hydrolytic enzymes. The cavity is closed off from the solvent by the second domain.[6] A representation of DhlA structure can be seen in Figure 4.1.

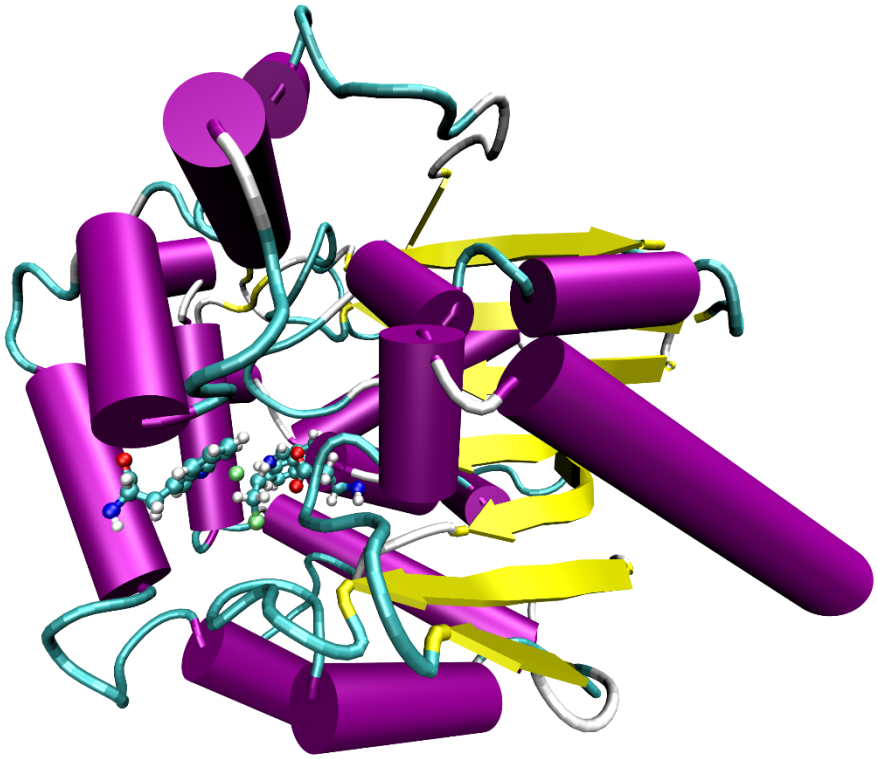


Figure 4.1: *Three-dimensional representation of the DhlA. Substrate has been displayed in balls and sticks format, α-helix in magenta and β-sheets in yellow.*

The reaction mechanism begins with the formation of the Michaelis complex, with the halogen atom from the substrate hydrogen bonded between two residues present in the cavity, Trp125 and Trp175. This step is followed by the nucleophilic attack of the carboxylate anion of Asp124, resulting in a covalently bound intermediate and a halide ion that remains hydrogen-bonded to the previous residues. This is the step that will be simulated in our project. Afterwards, the resulting intermediate is hydrolyzed by a water molecule activated by His289 as base, followed by

fast release of the alcohol product and finishing with the exit of the halide from the active site.[6]

System Definition

The Cartesian coordinates of the protein were taken from the X-ray crystallographic structure obtained from the Protein Data Bank (code 2DHC)[5]. The protonation state of each residue was determined at pH 7.0, employing the PropKa 3.1 program.[7-10] Once the protonation state of each titratable residue was determined, hydrogen atoms were added by means of the HBUILD tool included in the fDYNAMO package.[11] To neutralize the charge of the protein, 16 sodium ions were added so that the total charge was the same as the correspondent aqueous solution system (-1) because of the nucleophilic oxygen. The whole system was placed in a pre-equilibrated cubic box of water molecules of side 79.5 Å, deleting all those water molecules with oxygen atoms found at 2.8 Å from any non-hydrogen atom of the protein. For the system in aqueous solution, we placed DCE and acetate, which represented the substrate and the reacting fragment of Asp residue in the active site in a pre-equilibrated box of water molecules of side 55.8 Å, using the same procedure as in the enzymatic system.

Our systems were modelled using a quantum mechanics/molecular mechanics (QM/MM) computational scheme in which DCE and the side chain of residue Asp124 (acetate molecule for aqueous solution) are described using the PM3 semi-empirical Hamiltonian (see Figure 4.2).[12] To saturate the valence of the QM/MM frontier we used the link atom procedure previously explained in Chapter 2. It is important to note that

the PM3/MM Hamiltonian results in systematically overestimated energy barriers, but the geometries obtained for the RS and TS are good enough for reasonable estimations of Kinetic Isotope Effects (KIEs).[13]

The MM subsystem is defined in a similar way for every system in this project: water molecules are described with a flexible TIP3P potential;[14,15] while the rest of the enzyme, the counter-ions and any cofactor, if any, that was not included in the QM part are described by means of the all-atoms optimized potential for liquid simulation (OPLS).[16-18] The van der Waals parameters of the QM atoms are also taken from the OPLS potential, except those for the chlorine atoms, whose parameters are taken from Gao and Xia.[19]

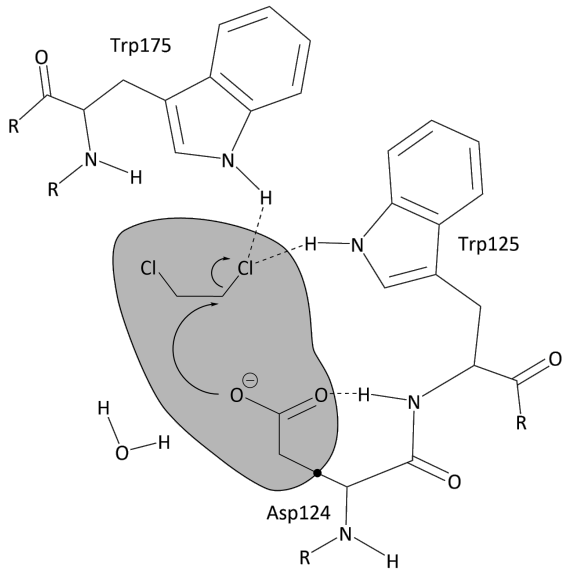


Figure 4.2: *Schematic representation of the S_N2 reaction catalyzed by DhlA. The shaded part corresponds to the QM region and the black dot represents the "link atom".*

After the preparation and definition of the system, we carried out a simulation of first equilibration molecular dynamic in the NVT ensemble at a reference temperature using the Langevin integrator. In this case, the two systems were equilibrated by means of 200 ps of molecular dynamics simulation at the reference temperature of 298 K with a time step of 1 fs and employing periodic boundary conditions. A cutoff radii switched between 12.5 and 15 Å was applied for all non-bonded of interactions. These conditions will be kept during all the MD simulations performed to these systems.

Results and Discussion

In this work we obtained the FESs as a function of two different coordinates, a solute coordinate (q) and a solvent coordinate, specifically the electrostatic coordinate (Ep). Then the FES can be expressed as:

$$W(q, Ep) = C' - kTln \cdot \\ \cdot \int \rho(\boldsymbol{x^N})\delta\Big(q(\boldsymbol{r^N}) - q_0)\Big)\delta(Ep(\boldsymbol{r^N}) - Ep_0)dx^N \tag{4.1}$$

where $\rho(\boldsymbol{x^N})$ is the probability density of finding the system at the configuration $\boldsymbol{x^N}$. In this case, q is the antisymmetric combination of the distances of the outgoing chloride and the incoming oxygen to the carbon atom.

$$q = d(ClC) - d(OC) \tag{4.2}$$

The electrostatic coordinate selected, on the other hand, was the antisymmetric combination of the electrostatic potential created by the environment on the outgoing chlorine atom and the incoming oxygen atom:

$$Ep = V_{CL}(\boldsymbol{r^N}) - V_O(\boldsymbol{r^N}) \tag{4.3}$$

The FESs were obtained using umbrella sampling, applying parabolic constraints to the solute and the solvent coordinates. The force constants used to keep the system at the reference values of the solute and solvent coordinates were 2,500 $kJ \cdot mol^{-1} \cdot \text{Å}^{-2}$ and 0.01$kJ^{-1} \cdot mol \cdot |e|^2$, which provided a good control of the coordinates.[20] To save computational cost, simulations were performed with any atom beyond 25 Å of dichloroethane frozen. A total of 5,454 simulation windows were employed to trace the FESs in the solution while for the enzyme 3,100 windows were needed. Each window consisted of 5 ps of equilibration and 50 ps of production.

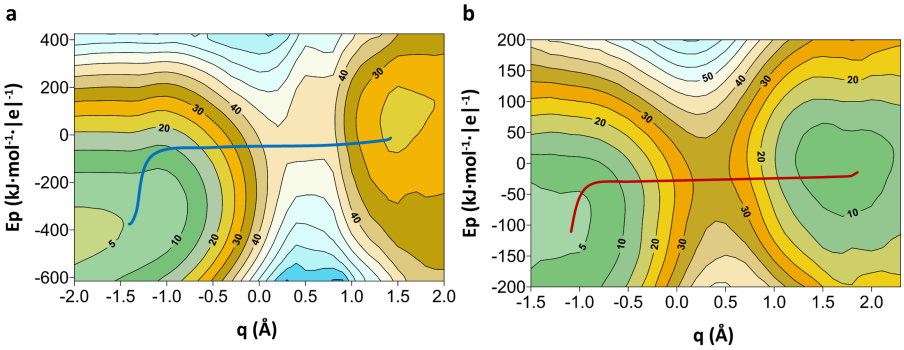


Figure 4.3: *Free Energy surfaces for the dehalogenation S_N2 reaction of dichloroethane for the uncatalyzed reaction* (a) *and in the active site of DhlA*(b)*. The isoenergetic free energy lines are drawn each 3 kcal* $\cdot mol^{-1}$

and the continuous blue and red lines represent the minimum free energy path (MFEP) on the FESs.

The free energy differences between the saddle points and the reactant minima located on the FESs are 27.4 and 37.1 $kcal \cdot mol^{-1}$ for the catalyzed and uncatalyzed processes, respectively. The activation free energy deduced from the experimental rate constant of the enzymatic reaction at 298 K is 15.3 $kcal \cdot mol^{-1}$ [21] and the barrier estimated for the process in aqueous solution is 26 $kcal \cdot mol^{-1}$.[22] This overestimation of the free energy barriers results from the use of the PM3 Hamiltonian,[13,23] but the geometries obtained for the RS and TS are good enough for reasonable estimations of KIEs.[13]

We then corrected this systematic error in the activation free energies by means of single-point calculations at higher theoretical levels. With this purpose we optimized ten TS structures starting from different configurations selected from the corresponding simulation window. After intrinsic reaction coordinate calculation, the energy barrier was obtained at the PM3/MM level and by means of single-point calculations at the M06-2X/6-311+G(2df,2p)/MM level.[24] The correction energy term was evaluated as the averaged difference between the semi-empirical and M06-2X energy barriers. The M06-2X corrected free energy barriers for the enzymatic and in-solution processes are 16.5 and 27.4 $kcal \cdot mol^{-1}$, respectively, in better quantitative agreement with the experimental values. In any case, the PM3/MM calculations provide a correct estimation of the catalytic effect, defined as the difference between the in-solution and the enzymatic free energy barriers. The PM3/MM difference is 9.7 $kcal \cdot mol^{-1}$, in good agreement with the difference derived from

the M06-2X values, 10.9 $kcal \cdot mol^{-1}$, and from the experimental values, 10.7 $kcal \cdot mol^{-1}$.

The FESs show noticeable differences between the reaction in the solution and in the enzyme. For the RS the protein structure provides a much more adequate environment for the progress of the reaction than does the solution. In solution, the RS stays at a value for the solvent coordinate of about -400 $kJ \cdot mol^{-1} \cdot |e|^{-1}$, while the enzymatic RS remains at about -100 $kJ \cdot mol^{-1} \cdot |e|^{-1}$. This latter value of the solvent coordinate is much closer to the value needed to reach the TS, at around -40 $kJ \cdot mol^{-1} \cdot |e|^{-1}$ for both environments. For the Michaelis complex the protein is already organized, from the electrostatic point of view, to favor the reaction, but in aqueous solution the environment needs a larger reorganization to facilitate the reaction.[25-27] Nevertheless, the protein structure does not behave as a rigid scaffold in which the reaction takes place. The reaction would be significantly more difficult in a frozen-protein environment in which the solvent coordinate remains unchanged from RS to TS.

To characterize the time evolution of the environment we computed the characteristic frequencies associated with the motion along the solvent coordinate in the two environments using force constants and effective masses deduced from the equipartition principle. The results shown in Table 4.1 demonstrate that the protein structure is stiffer than the structure of water, since the force constant obtained in the enzyme is about 4.2 times larger than that obtained in the solution. This is related to the existence of a network of covalent bonds in the enzymatic system.

Table 4.1: *Force constants (K_s in $kcal^{-1} \cdot mol \cdot |e|^2$), effective masses ($m_s$ in $kcal^{-1} \cdot mol \cdot |e|^2 \cdot s^2$) and characteristic frequencies (v_s in cm^{-1}) associated with the solvent coordinate (s) for the S_N2 reaction in solution and in DhlA.*

	Aqueous Sol.	**DhlA**
K_s	3.3×10^{-3}	1.4×10^{-2}
m_s	0.4×10^{-30}	2.3×10^{-30}
v_s	480	410

However, as stated above, the change needed in the solvent coordinate to reach the TS from the RS is much smaller in the enzyme than in the solution. The final result is that the work to be done on the solvent coordinate to reach the TS is significantly smaller in the enzyme than in the solution. According to the MFEP traced on the FESs, the free energy difference between the TS and the RS can be written approximately as:

$$\Delta G^{\ddagger} \approx \Delta G_s(Ep_{RS} \rightarrow Ep^{\ddagger}, q_{RS}) + \Delta G_q(Ep^{\ddagger}, q_{RS} \rightarrow q^{\ddagger}) \qquad (4.4)$$

where $\Delta G_s\left(Ep_{RS} \rightarrow Ep^{\ddagger}, q_{RS}\right)$ represents the work to be done on the solvent coordinate and $\Delta G_q\left(Ep^{\ddagger}, q_{RS} \rightarrow q^{\ddagger}\right)$ the work to be done on the solute coordinate.

The results presented in Table 4.2 show a large difference in the contribution to the free energy barrier of the solvent coordinate between the enzymatic system and the aqueous solution, being the free energy cost associated with the change along the solvent coordinate in the enzyme substantially smaller than that in the solution.

This difference represents 80% of the catalytic effect, since the difference between the contributions to the free energy barrier of the solute coordinate is much smaller.

Table 4.2: *Free energy barrier ($\Delta G^{\ddagger}$) and its respective contributions along the solvent (ΔG_s) and solute (ΔG_q) coordinates for the reaction in aqueous solution and DhlA.*

$(kcal \cdot mol^{-1})$	**Aqueous Sol.**	**DhlA**
$\Delta G^{\ddagger}$	38	28
ΔG_s	11	3
ΔG_{ξ}	27	25

Another important aspect to be analyzed is the timing between the solute and solvent motions. To characterize the time evolution of the environment we computed the characteristic frequencies associated with the motion along the solvent coordinate in the two environments, using force constants and effective masses deduced from the equipartition principle. These frequencies are provided in Table 4.1. As observed, both the force constant and the effective mass associated with the solvent coordinate are larger in the enzyme than in the solution. Both effects cancel out and, as a result, the frequency associated with the motion along the solvent coordinate in the enzyme ($410\ cm^{-1}$) is very similar to the value obtained in the solution ($480\ cm^{-1}$). These values essentially correspond to the reorientation of hydrogen-bond donors around the nucleophile and the leaving group, motions that occur in picoseconds or faster.[26] So, from the dynamical point of view there are no significant

differences in the participation of the environment during the reaction progress in aqueous solution or in DhlA.

We now proceed to compare the results obtained using both the solute and the solvent coordinates with those obtained assuming equilibrium solvation at any value of the solute coordinate. In the equilibrium assumption, the environment is at its free energy minimum at every value of the solute coordinate. The representation of the Equilibrium Free Energy Path (EFEP), i.e. the path obtained assuming equilibrium solvation along the solute coordinate, is presented in Figure 4.4 along with the MFEP.

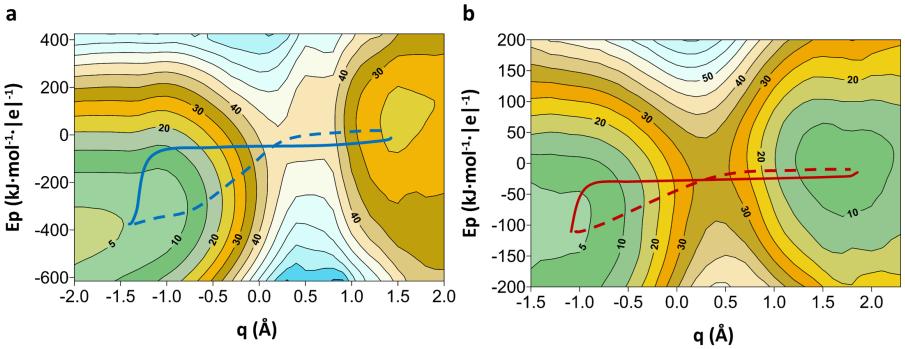


Figure 4.4: *Comparison between the MFEP and the EFEP on the FESs obtained in solution* (a) *and DhlA* (b).

It can be seen that the EFEP match the MFEP at the stationary points of the surface. Then 1D profiles traced along the solute coordinate provide almost the same activation free energies as those of 2D surfaces obtained as a function of the solute and solvent coordinates. However, although the free energy differences are correct, the equilibrium solvation approach is unable to describe properly the timing between the solute and solvent

coordinates. Effectively, as observed in Figure 4.4, in the equilibrium treatment one pulls along the solute coordinate and the solvent coordinate abruptly changes in the vicinity of the TS, but along the MFEPs, both in the solution and in the enzyme, solvent motions precede the changes along the solute coordinate. In any case, this limitation does not affect the estimation of reaction-rate constant, because this is determined mostly by the free energy difference between the TS and the RS.

Although the MFEPs and EFEPs coincide at the free energy minimum and at the saddle point, there is a small difference in the activation free energies estimated from the 2D (from the non-equilibrium pathway) and the 1D (or equilibrium pathway) treatments. The origin of this difference is in the definition of the TS ensemble obtained in each treatment. On the one hand the dividing surface defined in the non-equilibrium description contains the saddle point and goes through the ridges that separate the reactants and products valleys. On the other hand, in the equilibrium description, only the solute coordinate is employed and thus the dividing surface is defined simply as $q = q^{\ddagger}$ (see Figure 4.5a and 4.5b). The TS further along the non-equilibrium dividing surface is narrower than the TS further along the equilibrium dividing surface (see Figure 4.5c) and so the frequencies associated with the motion of the TS along the former dividing surface ($v_{Ep,q}$) are larger than those for the latter (v_q).

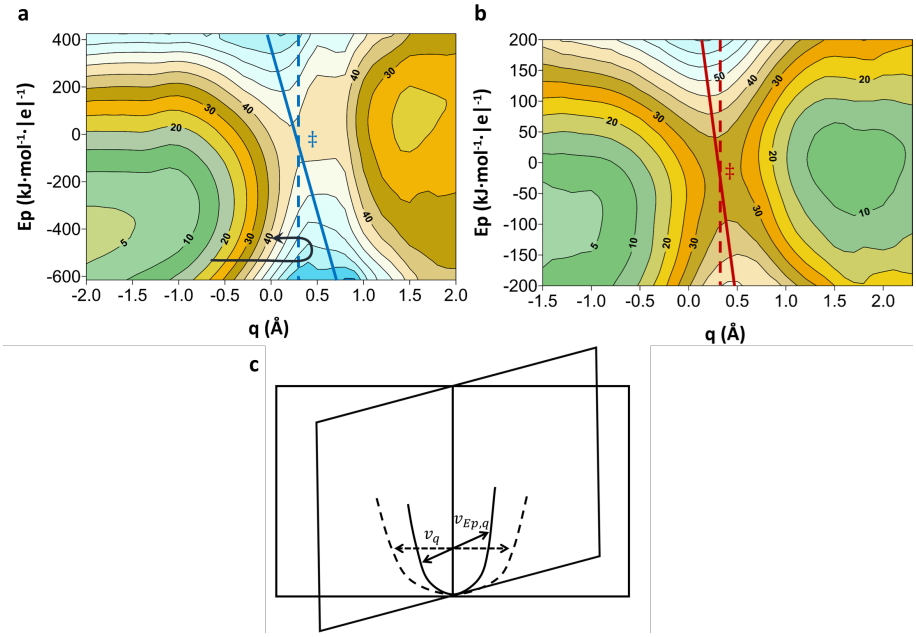


Figure 4.5: *Dividing surfaces obtained from equilibrium (dashed lines) and non-equilibrium pictures (continuous lines) for the uncatalyzed (*a*) and the catalyzed reaction (*b*). The arrow represents a hypothetical trajectory recrossing the equilibrium dividing surface. Schematic representation of the TS wells projected on the equilibrium and non-equilibrium dividing surfaces with their characteristic frequencies (*c*).*

This can be translated into an entropic difference in the TS ensembles and consequently in the activation free energies:[28]

$$\Delta\Delta G^{\ddagger} = \Delta G^{\ddagger}(Ep,q) - \Delta G^{\ddagger}(q) = RTln\frac{v_{Ep,q}}{v_q} \qquad (4.5)$$

From the FESs we could take the free energy profile along the dividing surfaces and calculate the frequencies both for the aqueous solution and for the enzyme. The ratios between them are about 1.3 and 1.1

respectively; consequently the $\Delta\Delta G^{\ddagger}$ at 298 K are about 0.2 and 0.1 $kcal \cdot mol^{-1}$. These free energy differences are below the statistical uncertainty of typical free energy simulation, which indicates that non-equilibrium effects make a very small contribution to the activation free energies.

Obviously, a smaller activation free energy is translated into a larger rate constant. This difference should then be compensated by the consideration of a transmission coefficient smaller than unity. The origin of this difference can be understood in terms of the differences between the dividing surfaces obtained in the non-equilibrium and the equilibrium descriptions on our reduced 2D model. The dividing surface obtained from the non-equilibrium treatment is so defined that any trajectory arriving at that surface from the reactant side will continue to the product region because the free energy continuously decreases in that direction, and so the transmission coefficient is equal to unity for that surface. However, using the equilibrium dividing surface, some trajectories that go from reactants to products find a free energy barrier after crossing this surface and they could return to the reactant side (see Figure 4.5); meaning that the transmission coefficient would be smaller than unity. With the inclusion of the different transmission coefficients for each definition, the rate constants obtained from the non-equilibrium and the equilibrium approaches would be the same.

$$\kappa(q) = e^{-\frac{\Delta G^{\ddagger}(Ep,q) - \Delta G^{\ddagger}(q)}{RT}} \tag{4.6}$$

Using the free energy differences given above, the transmission coefficients for the reaction in the solution and in the enzyme obtained using the solute coordinate as the distinguishing reaction coordinate are 0.8 and 0.9, respectively. Obviously, this procedure leads to a very crude estimate of the transmission coefficient because of the statistical errors associated with the free energies and of the non-explicit treatment of all the degrees of freedom. A more accurate estimation, based on the "positive flux" formulation,[29] shows that the transmission coefficients for this reaction are about 0.6 and 0.8 in the aqueous solution and in the enzyme, respectively.[26]

4.2. Catechol O-Methyltransferase

The enzyme catechol O-methyltransferase catalyzes the methyl transfer reaction from S-adenosylmethionine (SAM) to the nucleophilic hydroxylate oxygen atom of catechol.[30-32]

Catechol O-methyltransferase (COMT) is an important enzyme in several different areas of dopaminergic pharmacology that are directly linked to levodopa therapy, a precursor of catecholamines and an important substrate of COMT. COMT inhibitors increase the elimination half-life of levodopa and prolong its availability to the brain, making them a significant and beneficial adjunct to levodopa therapy in Parkinson's disease patients.[33,34]

The atomic structure of COMT from rat liver has been resolved with a resolution of 2.0 Å.[30] The active site of COMT includes the coenzyme SAM; which is connected with van der Waals interactions with Trp143, His142 and Met91 residues, and also hydrogen bonded to Ser119. In the active site of COMT there is a Mg^{2+} ion, which has a crucial role in the substrate binding.[30] The Mg^{2+} is coordinate to the side-chain oxygens from residues Asp141, Asp169, Asn179 and to both inhibitors hydroxyls, the sixth position of the octahedral magnesium coordination is occupied by a water molecule. The binding site for catechol is close to the surface of the protein without any part of it enclosing the substrate.[30] A schematic representation of COMT's structure can be seen in Figure 4.7

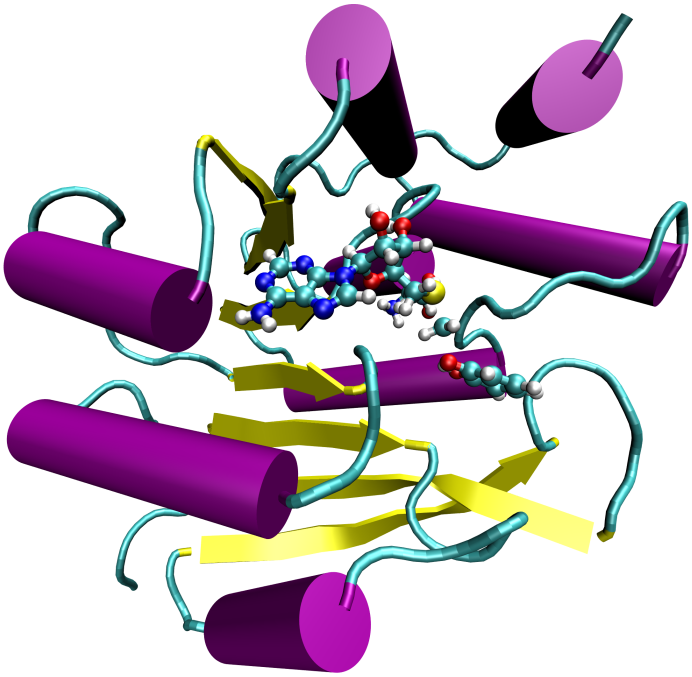


Figure 4.7: *Three-dimensional representation of the COMT. Substrate and cofactor have been displayed in balls and sticks format, α-helix in magenta and β-sheets in yellow.*

The first step of the catalytic mechanism is the binding of the Mg^{2+} ion. The SAM cofactor only binds after the enzyme is complexed with the Mg^{2+} ion and finally the substrate enters the enzyme.[35] During the chemical step, the methyl fragment is transferred from the cofactor to the substrate, as Figure 4.8 shows.

System Definition

The initial structure for the simulations carried out in this work was taken from the X-ray structure with Protein Data Bank code 1VID, which contains the ternary complex of the enzyme with the SAM cofactor and the inhibitor 3,5-dinitrocatechol.[30] Both nitro groups on the inhibitor

were removed to prepare the natural substrate, which also needs to be ionized by proton transfer to neutral Lys144.[36]

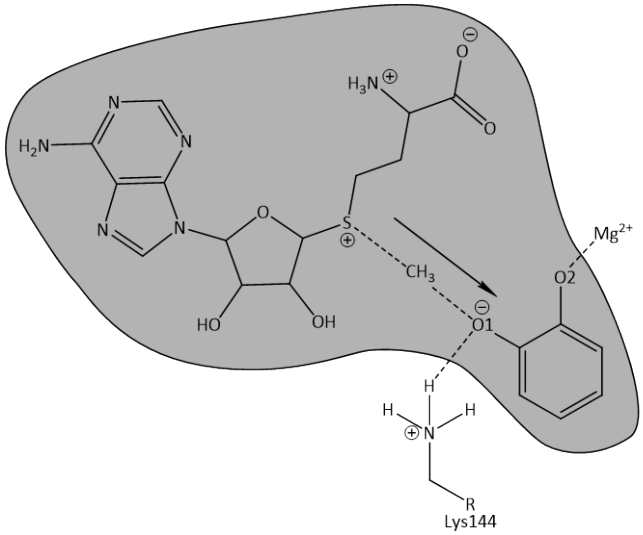


Figure 4.8: *Schematic representation of the methylation reaction mechanism catalyzed by COMT. The shaded part corresponds to the QM region.*

After applying the same procedure described in System Definition of section 4.1 with the DhlA system, the whole system was placed inside a preequilibrated cubic box of water molecules of side 79.5 Å and two sodium ions where added to the enzymatic system to keep it electroneutral. To study the reaction in solution, we followed the same procedure placing the whole cofactor SAM and the substrate in a preequilibrated water box of side 55.8 Å.

We used a QM/MM computational scheme in which SAM and catecholate (63 atoms) are described using the AM1 semiempirical Hamiltonian both in enzyme and in aqueous solution.[37] This QM level overestimates the free

energy barrier of the reaction but provides reasonable geometries for the stationary structures, with a slight underestimation of the distance between the carbon atoms of the transferred methyl group to the nucleophilic oxygen atom of catecholate.[38,39] The MM subsystem was formed by the enzyme, the magnesium cation, crystallization waters and solvating water molecules.

Results and Discussion

We obtained the FESs for this system in a similar way we did in section 4.1. In this case, the substrate coordinate (q) is the antisymmetric combination of the distances of the SAM sulphur atom (S) and catecholate oxygen (O), see Figure 4.8, to the carbon atom of the transferred methyl group (C):

$$q = d(SC) - d(O1C) \tag{4.7}$$

On the other hand, the solvent coordinate was the antisymmetric combination of the electrostatic potential created by the environment on the sulphur and oxygen atoms:

$$Ep = V_{O1}(\boldsymbol{r^N}) - V_S(\boldsymbol{r^N}) \tag{4.8}$$

The solvent coordinate changes from large positive values at the reactant state to smaller values at the product state due to the charge separation annihilation that occurs during the reaction.

It must be noticed that, by construction, the solute and the solvent coordinates are not orthogonal, as they involve the coordinates of some

common atoms (the methyl donor and acceptor). However, we checked that at the TSs, these coordinates are on average almost orthogonal. The averaged angles between them, determined from the gradients in the mass-weighted Cartesian space, were 90.3 ± 0.5 and 95.4 ± 0.4 degrees in aqueous solution and in the enzyme respectively.

The force constants used to keep the system at the reference values are the same as those used for the DhlA system. A total of 2840 simulation windows were employed to trace the FES in the aqueous solution, while for the enzyme 1420 windows were needed. The difference comes from the fact that the reaction takes place spanning a wider range of values of the solvent coordinate in the aqueous solution than in the enzyme, as shown in Figure 4.9. Each simulation window consisted of 5 ps of equilibration and 45 ps of production.

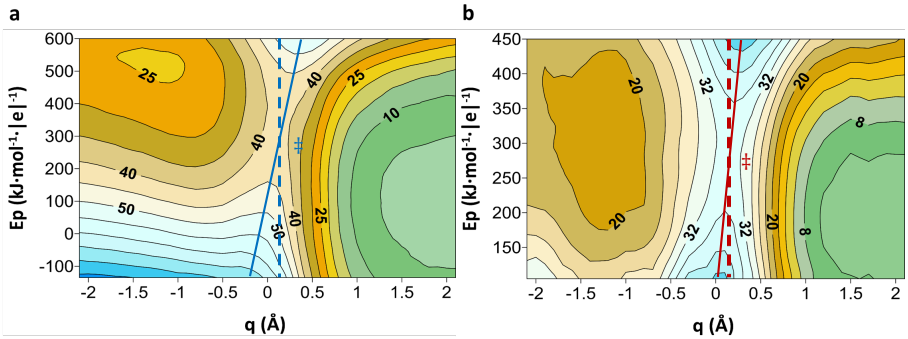


Figure 4.9.: *Free energy surfaces corresponding to the aqueous solution reaction (*a*) and the enzymatic reaction (*b*). Isocontour lines are drawn each 5 kcal · mol*$^{-1}$*. Continuous lines represent the TS ensembles obtained using the non-equilibrium definition (*$TS_w(q, Ep)$*,* $TS_e(q, Ep)$*). While dashed lines represent the TS ensemble corresponding to the equilibrium solvation (*$TS_w(q)$*,* $TS_e(q)$*).*

The approximate positions of the TS ensembles were first estimated from the corresponding FESs. Then 100 ps long QM/MM simulations were performed with a biasing potential added to restrain the system around the saddle point. Assuming that the region of the FES accessible to such a simulation can be well described within the harmonic approximation, we calculated the precise position of the saddle point ($\boldsymbol{\xi^{TS}}$) on the FESs as:[40]

$$\boldsymbol{\xi^{TS}} = \boldsymbol{H}^{-1}(k_B \cdot T \cdot \boldsymbol{C}^{-1} \cdot \boldsymbol{\xi}(\boldsymbol{R})^{\boldsymbol{b}} - \boldsymbol{K} \cdot \boldsymbol{\xi^0}) \quad (4.9)$$

where $\boldsymbol{H}$ is the 2 × 2 Hessian matrix associated with the 2D-FES, $\boldsymbol{C}$ the covariance matrix, $\boldsymbol{K}$ the diagonal matrix of force constants, $\langle \boldsymbol{\xi}(\boldsymbol{R}) \rangle^{\boldsymbol{b}}$ the average position over the biased simulation and $\boldsymbol{\xi^0}$ the reference position of the bias. The Hessian matrix is obtained as.[40]

$$\boldsymbol{H} = k_B \cdot T \cdot \boldsymbol{C}^{-1} - \boldsymbol{K} \quad (4.10)$$

The saddle points in both FESs are found at very similar values of the solute and solvent coordinates. In the enzyme this saddle point is located at $q_{Enz}^{TS} = 0.13$ Å and $Ep_{Enz}^{TS} = 268\ kJ \cdot mol^{-1}|e|^{-1}$ while in aqueous solution the values found for the two coordinates are $q_{Wat}^{TS} = 0.08$ Å and $Ep_{Wat}^{TS} = 326\ kJ \cdot mol^{-1}|e|^{-1}$. Interestingly, while the solute and solvent coordinates take very similar values at the TS in solution and in the enzyme, large differences appear at the reactant regions. In the enzyme, the value of the solvent coordinate in this region is $Ep_{Enz}^{RS} \approx 356\ kJ \cdot mol^{-1}|e|^{-1}$, close to the value found for the TS. However, the value of the solvent coordinate in solution is $Ep_{Wat}^{RS} \approx 585\ kJ \cdot mol^{-1}|e|^{-1}$.

The different behavior of the environment in the catalyzed and non-catalyzed reaction illustrates the concept of enzyme preorganization.[41,42] As in DhlA, the electrostatic properties of the active site at the Michaelis complex of COMT are already close to those needed to reach the reaction TS. As it has been seen, the larger reorganization observed in solution implies a free energy penalty reflected in the larger barrier for the uncatalyzed reaction in solution.[43]

Having calculated the Hessian matrices and TS positions, a biasing potential restraining the simulation to sample the TS ensemble can be constructed as follows:

$$V_{umb}^{TS}\left(\boldsymbol{\xi}(\boldsymbol{R})\right) = \frac{K^{TS}}{2}\left((\boldsymbol{\xi}(\boldsymbol{R}) - \boldsymbol{\xi}^{\boldsymbol{TS}}) \cdot \boldsymbol{v}_{\boldsymbol{H}}\right)^2 \qquad (4.11)$$

where K^{TS} is the force constant, which must be large enough to keep the system close to the TS, and $\boldsymbol{v}_{\boldsymbol{H}}$ is the eigenvector of the Hessian that corresponds to the negative eigenvalue. With this bias, sets of 100 possible transition structures were obtained from MD simulations, saving those structures in which $V^{TS} < 0.5k_BT$ and that were separated by at least 1 ps of simulation. From each selected structure, 50 trajectories with random initial velocities taken from a Maxwell–Boltzmann distribution at T = 300 K were integrated both forward and backward in time in the NVE ensemble using the velocity Verlet integrator to obtain the probability to commit to the reactants and products basins.

According to the average reactive trajectories presented in Figure 4.10, most of the solvent reorganization takes place in the first stages of the reaction. The timing for the evolution of the solute and solvent

coordinates is clearer for the reaction in the aqueous solution, where the change in the solvent coordinate is substantially larger than in the enzyme.

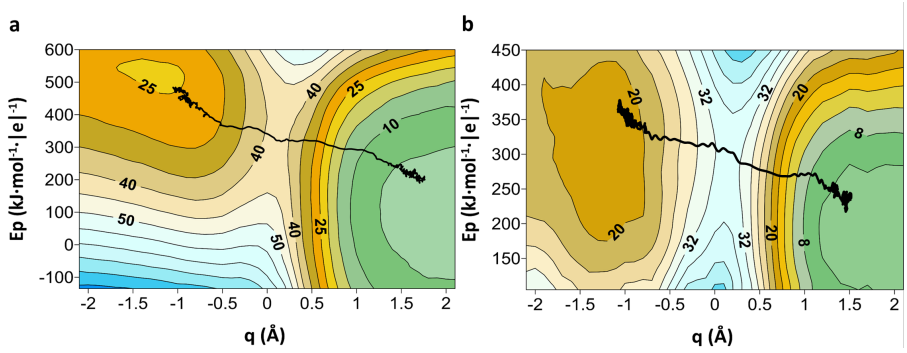


Figure 4.10: *Average reactive trajectories projected on the FES are represented with a black line.*

This fact is also supported by the calculated average frequencies, from the free energy landscape and effective masses deduced from the equipartition principle, for the solvent coordinate in the aqueous solution and in the enzyme (235 and 225 cm^{-1}), which are much smaller than the average frequency associated to the solute coordinate, about 750 and 770 cm^{-1}in the solution and in the enzyme, respectively.

However, one must take into account that *Ep* is a collective coordinate that involves many fundamental motions or fluctuations of the environment, from fast stretching motions to slow conformational or rotational motions. Looking for a better understanding of the participation of these motions on the *Ep* coordinate, we have computed both the autocorrelation function and the Fourier spectra, in solution and in COMT, of the solvent coordinate from 50 ps trajectories starting at the

TS and in those where the solute coordinate was kept frozen at $q = q^{TS}$ using a Lagrange-multiplier-based SHAKE algorithm.[44]

The time evolution of the autocorrelation function shows a slow decay in the picosecond time scale and also a fine structure reflecting the participation of faster components (see Figure 4.11). Interestingly, a slower decay is observed in the solution compared to the one in the enzyme. While in principle much slower motions than in the solution could happen in the enzyme (associated for example with global conformational changes), the participation of slow motions in the electrostatic coordinate seems to be more important in the solution.

The Fourier spectra of the solvent coordinate shows that the electrostatic coordinate is mainly the result of slow motions. However, in the solution it is observed a small fast contribution around 3600 cm^{-1} due to the O-H stretching of water molecules. A larger signal is observed ($\sim$3400 cm^{-1}) in the enzyme assigned to the N–H stretching of Lys144 , a positively charged residue that is hydrogen-bonded to the nucleophilic atom (see Figure 4.11). Some low intensity signals appear in both environments in the region between 1400 and 1800 cm^{-1}, corresponding to heavy atoms stretching in the protein and water molecule bending. However, the most intense signals are found below 50 cm^{-1}, which correspond to motions taking place on the picosecond time scale. Only those components of the solvent coordinate presenting higher frequencies could couple to the solute coordinate during the passage through the dividing surface, while the slower components must precede the motion along the solute. The enzyme shows more important participation of motions in the 100–500

cm^{-1} range than the aqueous solution, while in the aqueous solution some intense signals appear in the 600–700 cm^{-1} range, a frequency similar to that of the solute coordinate, suggesting a more significant participation of the environmental motions in the reaction coordinate at the TS. Hydrogen bonds and electrostatic interactions between the environment and the reacting system appear in these regions of the spectra.

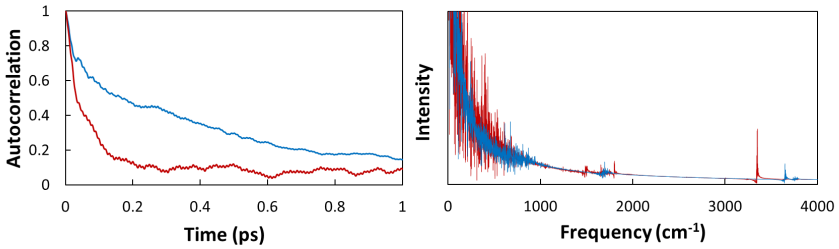


Figure 4.11: *Autocorrelation function (left) and Fourier transform (right) in the Electrostatic potential coordinate determined from simulations at* $q = q^{TS}$ *for the system in the aqueous solution (blue line) and in COMT (red line).*

Our interest will be focused now in the characterization of TS ensembles, in both environments, defined under the equilibrium solvation description ($TS(q)$) and using the non-equilibrium description ($TS(q, Ep)$). It can be observed that geometrical parameters shown in Table 4.3 are statistically indistinguishable for both definitions of the TS. The averaged geometrical parameters corresponding to the broken and formed bonds, the donor–acceptor distance and the attacking angle are similar, with differences below the standard deviations. Slightly larger changes are observed between the TS ensembles in the aqueous solution, in particular for the methyl attacking angle (S–C–O1), although the

dispersion associated with the structures of the TS ensembles prevent any quantitative conclusion from this observation. We also provide some distances characterizing the interaction of the chemical system to the environment. In the enzyme, the distance of the hydroxyl oxygen atom of catechol (O2) to the magnesium ion present in the enzymatic active site and the distance of the nucleophilic oxygen O1 to the ammonium group of Lys144 are equivalent in both TS ensembles. In the solution, the average distance of the O1 atom to the closest water molecule (O1–O_w) is also equivalent in both ensembles within their standard deviation.

Table 4.3: *Average geometrical parameters (distances (D) in Å, angles (<) in degrees) and transmission coefficients (κ) obtained for the TS ensembles characterized using the solute (q) or the solute and solvent coordinates (q, Ep).*

	Aqueous Solution		**COMT**	
	$TS_w(q)$	$TS_w(q,EP)$	$TS_e(q)$	$TS_e(q,EP)$
D(S-C)	2.12 ± 0.04	2.12 ± 0.04	2.17 ± 0.05	2.16 ± 0.05
D(C-O1)	2.03 ± 0.04	2.03 ± 0.05	1.99 ± 0.05	1.98 ± 0.05
D(S-O1)	4.14 ± 0.08	4.11 ± 0.08	4.12 ± 0.09	4.11 ± 0.08
<(S-C-O1)	172 ± 4	167 ± 6	166 ± 5	166 ± 6
D(Mg-O2)			2.23 ± 0.09	2.24 ± 0.10
D(O1-N Lys144)			2.74 ± 0.09	2.75 ± 0.09
D(O1-O_w)	2.76 ± 0.07	2.73 ± 0.08		
κ	0.49 ± 0.05	0.70 ± 0.05	0.76 ± 0.05	0.72 ± 0.05

It is interesting to note that our results do not show any evidence for a significant difference in the distances between the methyl donor and acceptor in the enzymatic TS with respect to that of the aqueous solution. The experimental observation of much more inverse secondary α-deuterium KIEs for the COMT catalyzed reaction compared to the uncatalyzed methyl transfer was originally interpreted as the consequence of a tighter transition state in the enzymatic reaction, with significantly smaller donor–acceptor distance.[45] However, QM/MM calculations based on equilibrium simulations along the solute coordinate reproduced the experimental observation without evidence for compression.[46,47] While it could be argued that the sampling resulting from those simulations could be biased to produce similar results irrespectively of the environment, the new simulations presented here show that the averaged distances are still the same, within their standard deviations, in both environments. The explicit inclusion of environmental motions in the definition of the TS ensemble does not affect this conclusion, suggesting that the inverse secondary α-deuterium KIES in COMT[48-50] can be explained by the preferential equilibrium solvation provided by the environment.[46]

TS ensembles were also characterized analyzing a set of 5000 trajectories initiated from TS structures selected for each of the four ensembles (50 trajectories started from 100 TS structures for each ensemble, using different velocities taken from a Maxwell–Boltzmann distribution). From these trajectories, assuming that all have the same probability, we computed the transmission coefficient as described in Equation 2.92 in Chapter 2. The values of the transmission coefficient for the four TS

ensembles and their standard deviation are provided in Table 4.3. The values of the transmission coefficient are similar to those obtained in previous studies of the same system using slightly different simulation conditions.[51,52] It is interesting to note that the transmission coefficients calculated from the enzymatic TS ensembles optimized along one ($TS_e(q)$) and two ($TS_e(q, Ep)$) coordinates are statistically identical, while the transmission coefficient in water increases when passing from the $TS_w(q)$ to the $TS_w(q, Ep)$. This effect illustrates the fact that the participation of solvent motions during barrier crossing is significantly more important in the aqueous solution than in the enzyme. Moreover, when solvent motions are incorporated into the definition of the TS ensemble, the transmission coefficients are very similar in both environments. This runs against the proposals about a larger role of dynamical contributions during barrier crossing in enzymatic reactions.[53]

A more stringent test about the quality of the TS ensembles is provided by the analysis of the committor histograms. The histograms presented in Figure 4.12 provide the probability that free trajectories initiated from a given TS structure reach the basin of the products before the basin of the reactants. For each of the 100 structures selected for the four TS ensembles, the probability of ending in products for each structure was determined from the set of 50 trajectories. If the TS ensemble is correctly defined, the committor histogram should be a unimodal distribution peaked around 0.5.

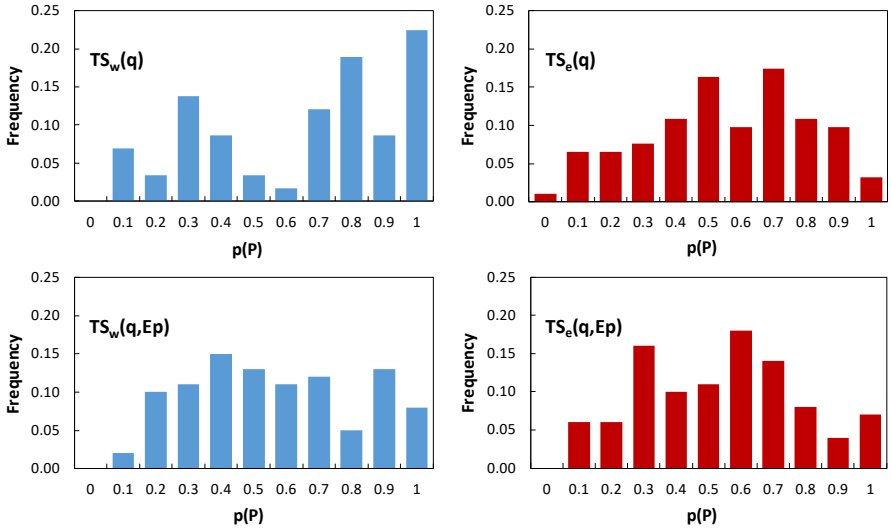


Figure 4.12: *Committor histograms obtained for the four TS ensembles analyzed from the probabilities that trajectories initiated from a given trajectory commit to products ($p(P)$). The graphs on the left correspond to the enzymatic reaction and those on the right to the aqueous solution reaction (one-dimensional and two-dimensional definitions up and down, respectively).*

The broadness observed for the committor distributions presented in Figure 4.12 can be due to at least three factors:[54]

1. An incomplete optimization of the reaction coordinate (the gradient of the committor function around the TS is usually high).

2. An insufficient sampling in terms of initial structures and/or trajectories.

3. Nonlinear effects in the dynamics.

According to Figure 4.12, the explicit consideration of the solvent coordinate in the definition of the TS surface in the aqueous solution

clearly improves the distribution of probabilities to commit to the product basins. Then a more rigorous picture of the TS can be obtained when the solvent coordinate is explicitly considered in the aqueous solution. The TS ensemble obtained using exclusively the solute coordinate shows a bimodal distribution that indicates the presence of a small barrier along the solvent coordinate. This barrier can be estimated from the ratio of the transmission coefficients obtained including an optionally solvent coordinate.

$$\Delta G^{\ddagger}(q, Ep) - \Delta G^{\ddagger}(q) = RT \cdot ln \frac{\kappa(q, Ep)}{\kappa(q)} \qquad (4.12)$$

This ratio equals 1.4 and can be translated into a free energy difference of only 0.2 $kcal \cdot mol^{-1}$ at 298 K, below the expected statistical error in the evaluation of activation free energies.

In the enzyme, the consideration of the solvent coordinate in the TS definition moderately improves the committor histogram. This result correlates with the changes observed in the transmission coefficients given in Table 4.3, which are almost equivalent in the two enzymatic TS ensembles. Participation of environmental motions in the reaction coordinate at the TS seems to be less important than in the solution. The larger reorganization needed in the solution to reach the TS, reflected in the larger variation observed for the electrostatic coordinate along the reaction (see Figure 4.10), results in a larger coupling between both coordinates in this environment. It must be emphasized that the solvent coordinate is composed of both fast and slow motions and only those motions presenting time scales similar to those of the solute coordinate

can couple during barrier crossing. Slower motions must precede the changes in the solute coordinate, which is reflected in the average reactive trajectory presented in Figure 4.10.

4.3. Dihydrofolate Reductase

Dihydrofolate Reductase (DHFR) takes part in the cellular metabolic cycle both in eukaryote and prokaryote cells.[55,56] It catalyzes the reduction of the 7,8-dihydrofolate (DHF) to 5,6,7,8-tetrahydrofolate (THF) using the NADPH as cofactor. At the biological level DHFR keeps the reserve of essential cofactors for other metabolic reactions, as THF.[57] This cofactor participates in the synthesis of thiamine nucleotide as a carbon and electron donor giving DHF as a product. In rapidly reproducing cells (like cancer cells) the synthesis of DNA is very high, so there is a big demand for thiamine nucleotide. Thus, the inhibition of DHFR is a therapeutic objective of various drugs aiming to reduce the proliferation of cancer cells.[58,59]

DHFR is a small enzyme consisting in only one monomer. Many studies have focused on the DHFR from *Escherichia coli* (EcDHFR)[53,57,60-84] , which is formed by a single chain and whose crystal structures have been solved at 1.7Å,[53,69,85] showing a complex secondary structure including several β sheets forming a hydrophobic layer around the active site. One of the loops forming this layer is the loop formed by residues 9 to 24, called the Met20 loop, which undergoes large scale millisecond time scale conformational changes along the enzymatic process and controls the entrance and departure of the substrate and cofactor.[57,65-67] This loop can adopt two conformations: the closed conformation, found in the holo-enzyme and the Michaelis complex, and the occluded conformation that follows the hydride transfer and prevents the nicotinamide ring of the cofactor NADP(H) from entering the active site.[57] This loop has been the

subject of several site-directed mutagenesis experiments aiming to understand the relevance of its flexibility.[53,60-64,66,70,86]

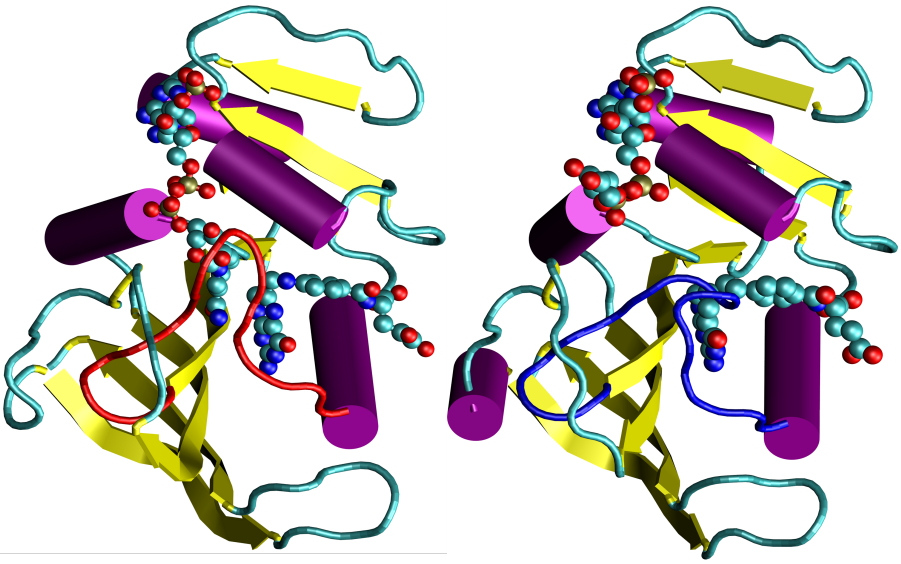


Figure 4.13: *Three-dimensional representations of DHFR. The Met20 loop is marked in red for the closed conformation (left) and blue for the occluded conformation (right). Substrate and cofactor are displayed in balls and sticks format.*

The first step of the mechanism, after the substrate and cofactor have entered the enzyme, is to create a positive charge density on C6 (see atom numeration in figure 4.14) to promote the hydride transfer from the cofactor. This can be achieved protonating N5 previously, as proposed by Huennekens and Scrimgeour in 1964.[87] After the protonation, the hydride transfer happens from the cofactor to the C6 atom. The calculations done over this system will be centered in the hydride transfer step to C6, which is the rate limiting step.[81]

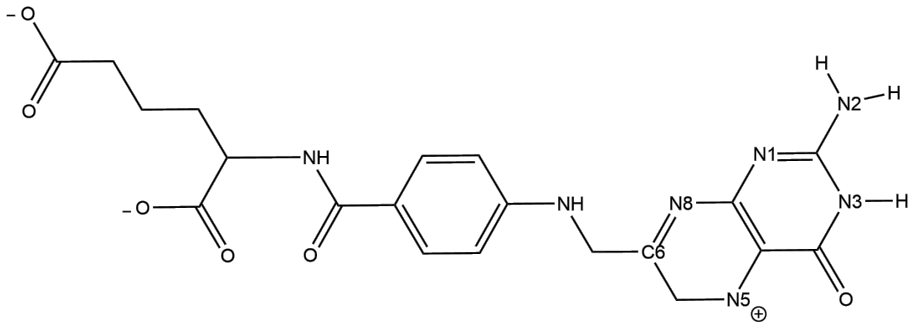


Figure 4.14: *Schematic representation of the 7,8-dihydrofolate structure.*

Dihydrofolate reductase (DHFR) has often been used in studies testing the relationship between protein motions and catalysis.[65,70,88-93] In particular the EcDHFR-N23PP/S148A mutation developed some controversy over the role of protein motions in hydrogen transfer reactions.[53,63,68] In the EcDHFR-N23PP/S148A mutant the Met20 loop is unable to adopt the occluded conformation due to the absence of the crucial hydrogen bonds needed to do so. Moreover, millisecond to microsecond time scale motions observed in the Met20 loop of wild type EcDHFR are lost in the variant.[53] This "dynamic knockout" displayed reduced hydride transfer rate constants, and it was proposed that the protein motions lost in EcDHFR-N23PP/S148A are involved in promoting hydride transfer in wild type EcDHFR.[53] However, theoretical studies based on the empirical valence bond approach suggested that the reduction in the hydride transfer rate constant in EcDHFR-N23PP/S148A was due to effects on the electrostatic preorganization and consequently the free energy of reorganization within the active site.[63]

Experimental Results

The calculations presented here are part of a collaboration project that included both theoretical calculations and experimental results.[62] In the experimental part the variant of Escherichia coli dihydrofolate reductase, EcDHFR-N23PP/S148A, was prepared together with its "heavy" version, where all the ^{14}N, ^{12}C and non-exchangeable ^{1}H atoms were isotopically substituted by their heavier version (^{15}N, ^{13}C, ^{2}H). Then the kinetic properties of both versions light (L) and heavy (H) were measured as presented in Table 4.4. The enzymatic KIES refers to the ratio between the rate constants of the L and H versions of the enzyme. The experimental data for the EcDHFR wild type, in its light and heavy versions, was obtained from ref.[94]

Table 4.4: *Experimentally Determined Rate Constants (k) and Enzyme KIEs (Enz. KIE) for Hydride Transfer in Light and Heavy EcDHFR-N23PP/S148A at 25 °C and activation Parameters from Fitting the Experimental Data to the Arrhenius Equation: Activation energy (E_a) and Arrhenius prefactor (A). Data for light and heavy wild type EcDHFR are from ref. 94.*

	$k(s^{-1})$	Enz. KIE	E_a $(kJ \cdot mol)$	$A(s^{-1})$
L EcDHFR	178.2 ± 4.7	1.10 ± 0.03	31.84 ± 0.69	(6.42 ± 0.81) × 10^7
H EcDHFR	151.6 ± 4.2		26.05 ± 1.45	(5.98 ± 0.11) × 10^6
L EcDHFR-N23PP/S148A	40.32 ± 0.79	1.33 ± 0.02	27.14 ± 0.16	(2.17 ± 0.14) × 10^6
H EcDHFR-N23PP/S148A	30.41 ± 0.80		25.62 ± 0.27	(8.63 ± 0.75) × 10^5

The results show an increase in the enzyme KIE from the wild type to the mutant. There is a reduction in the activation energies from the light to the heavy enzyme and a reduction in the Arrhenius prefactors and hence a less favorable activation entropy. The same situation has also been observed in the reaction catalyzed by the wild type EcDHFR.[94] The reduced rate constants for hydride transfer in EcDHFR-N23PP/S148A relative to EcDHFR have also been shown to be the result of less favorable activation entropy in the variant, rather than increased activation enthalpies.[68]

System Definition

The initial coordinates were taken from the X-ray crystal structures of the enzyme together with the substrate and the cofactor, using PDB files 3QL3[53] for the wild type and 3QL0[53] for the mutant. Heavy EcDHFR and EcDHFR-N23PP/S148A were prepared by modifying the masses of all ^{14}N, ^{12}C and non-exchangeable ^{1}H atoms to those of ^{15}N, ^{13}C, ^{2}H. The ratio between the masses of the simulated heavy and light variant enzymes was 1.10987, similar to the experimentally observed molecular weight increase.

After all protein versions were prepared with the same procedure described in section 4.1 and 4.2, 13, sodium cations were added in order to equilibrate the total charge of the system, with their initial positions selected according to the protein's electrostatic potential. The systems were then placed inside a pre-equilibrated cubic box of water molecules (65.2 Å sides) centered on the substrate-cofactor pair.

The uncatalyzed reaction in the aqueous solution was also studied following the same protocol, except that the substrate and the cofactor were first optimized together in gas phase using the MP2/6-31+G(d,p) ab initio method within the Gaussian09 package.[95] The optimized structure and 5 sodium ions (in order to compensate the total charge of the system) were then placed inside of a cubic box of water molecules (60.2 Å sides) centered on the substrate-cofactor pair.

In the QM/MM hybrid description, the protein, the ions, the water molecules and part of the cofactor and substrate formed the MM subsystem. The QM subsystem was formed by parts of the cofactor and the substrate that are directly involved in the chemical process, comprising 29 atoms from the NADPH coenzyme and 38 atoms from the protonated substrate (see Figure 4.15). To saturate the valence of the QM/MM frontier the link atom procedure was used. Therefore, the QM part was formed by 67 atoms plus 2 link atoms. The QM subsystem was treated by the semiempirical Hamiltonian AM1[37] as modified by Major et al. (AM1-SRP) to provide a more accurate description.[96]

Cutoffs for the nonbonding interactions were applied using a switching function within a radius range of 13.0 to 9.0 Å. Periodic boundary conditions were employed in all simulations. After full minimization and heating dynamics, the systems were equilibrated at 300 K and with a time step of 0.5 fs, due to the nature of the chemical step involving a hydrogen transfer over a total of 150 ps.

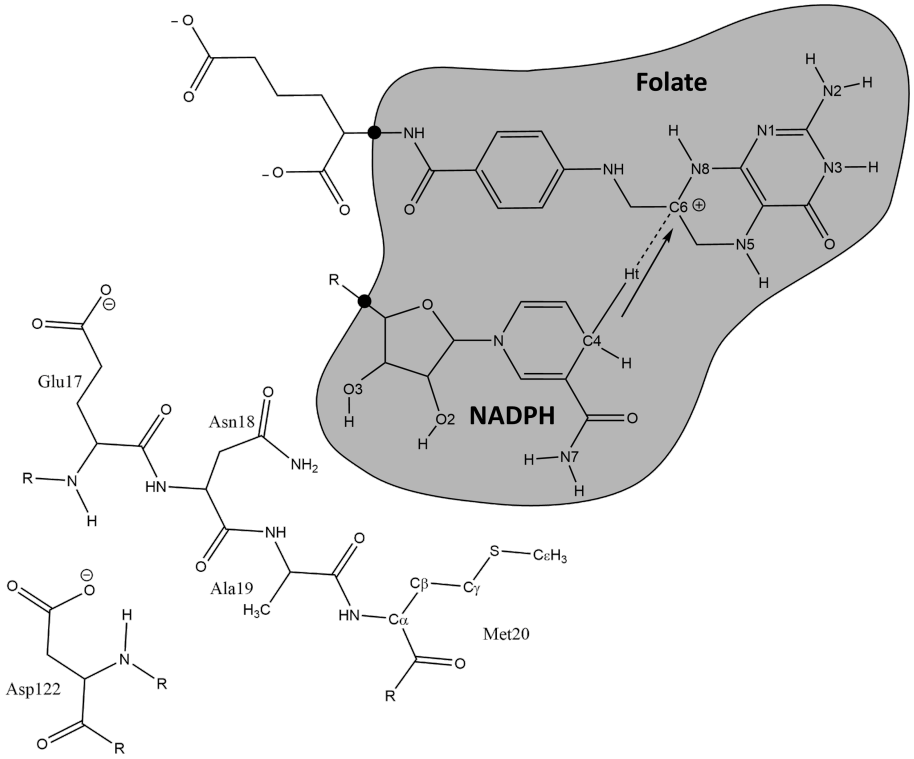


Figure 4.15: *Schematic representation of the hydrogen transfer reaction mechanism catalyzed by DHFR. The shaded part corresponds to the QM region and the black dots represent the "link atoms". Some important residues are also depicted.*

Results and Discussion

One-dimensional PMFs were computed for the EcDHFR-N23PP/S148A using the antisymmetric combination of distances ($q = d(C4H) - d(C6H)$, atom numbering from Figure 4.15), describing the hydride transfer as the reaction coordinate and using a force constant of 2500 $kJ \cdot mol^{-1} \cdot \text{Å}^{-2}$ to keep the system at the reference value. The probability distributions obtained from MD simulations consisted of 53 windows of 20 ps of relaxation and 40 ps of production, with a time step of 0.5 fs. Four

additional PMFs were computed to check the robustness of our method and to get averaged free energy values (see Figure 4.16). The starting structures were selected from snapshots of a long QM/MM-MD simulation with the reaction coordinate restrained to the value obtained for the TS of the first PMF.

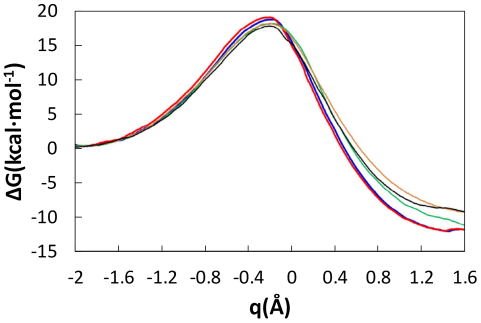


Figure 4.16: *PMF obtained from five different structures of the TS selected and optimized from the constrained MD simulation performed at the initial TS structure of EcDHFR-N23PP/S148A.*

The results show very small deviations between the profiles, and between the averaged structures of the three states involved in the reaction. From these PMFs, the classical activation free energy barrier obtained for EcDHFR-N23PP/S148A is $17.5 \pm 0.6\, kcal \cdot mol^{-1}$. A slightly higher dispersion than that of the wild type ($15.8 \pm 0.4\, kcal \cdot mol^{-1}$) was observed,[94] but the result is also in agreement with experimental data.[68,94]

Selected geometries of the RS and TS were used as starting points to run 2 ns of QM/MM-MD simulation with the reaction coordinate restrained to the corresponding values in order to investigate the structural properties of the RS and TS in more detail (see Table 4.5).

Table 4.5: *Key averaged structural parameters of the RS and TS. Distances (D) are in Å, angles (<) and dihedral angles (D<) in degrees. Data for light and heavy wild type EcDHFR are from ref. 94.*

	EcDHFR		EcDHFR-N23PP/S148A	
	RS	**TS**	**RS**	**TS**
<(C4-Ht-C6)	141 ± 15	163 ±7	144 ± 16	164 ± 7
D<(C4-Ht-C6-Ht)	-1.9 ± 0.4	-0.18 ± 0.04	-1.9 ± 0.4	-0.24 ± 0.04
D(C4-C6)	3.9 ± 0.3	2.63 ± 0.06	3.9 ± 0.3	2.66 ± 0.06
D(C4-Ht)	1.09 ± 0.03	1.24 ± 0.03	1.09 ± 0.03	1.22 ± 0.03
D(C6-H)t	2.9 ± 0.4	1.42 ± 0.04	3.1 ± 0.4	1.46 ± 0.04
D(N7-SMet20)	4.6 ± 0.7	2.6 ± 0.3	4.3 ± 0.4	3.0 ± 0.4
D(O3-NAsn18)	3.8 ± 0.6	4.1 ± 0.6	5.3 ± 0.5	4.0 ± 0.5
D(O2-OAla19)	3.2 ± 0.4	3.2 ± 0.5	3.4 ± 0.6	3.9 ± 0.6
D(O2-NAla19)	3.1 ± 0.2	3.0 ± 0.2	3.2 ± 0.4	3.3 ± 0.4
D<(Ca-Cb-Cg-SMet20)	166 ± 20	177 ± 13	175 ± 14	179 ± 16
D<(Ca-Cb-Cg-SMet20)	-83 ± 19	-75 ± 15	98 ± 40	-83 ± 28
D(HNGlu17-OAsp122)	4.0 ± 0.6	4.0 ± 0.7	3.2 ± 0.4	2.4 ± 0.6

In the case of EcDHFR-N23PP/S148A, the mutations alter some protein–protein and protein–cofactor interactions established by residues belonging to the Met20 loop at the reaction TS. In particular, the hydrogen bond distance between the Glu17 and Asp122 distance at the

TS of EcDHFR-N23PP/S148A (see Fig 4.15) is significantly shorter than in the wild type. The fact that this hydrogen bond is strongly formed only in the TS of EcDHFR-N23PP/S148A could contribute to the smaller enthalpic and larger entropic barrier found experimentally for the variant.[68] Other changes in the interactions established by the Met20 loop with the cofactor or the substrate are observed upon mutation. Met20 has a reduced capability to form a S-HN hydrogen bond with the amide group of the cofactor at the TS in the variant. From an electronic point of view, the formation of this hydrogen bond favors hydride transfer from the cofactor to the substrate.[97] Finally, other residues of the Met20 loop such as Asn18 and Ala19 are farther from the cofactor at the TS of the variant than in wild type EcDHFR, confirming the disruption of stabilizing interactions established by this loop in the TS by mutating the enzyme. Differences in the equilibrium fluctuations of the protein residues between wild type and variant were analyzed by means of the root-mean-squared fluctuation (RMSF) of all residues of the proteins. The differences observed between EcDHFR and EcDHFR-N23PP/S148A are never larger than 0.5 Å in absolute value as shown in Figure 4.17. Thus, both proteins show similar flexibility on the time scale relevant to catalysis.

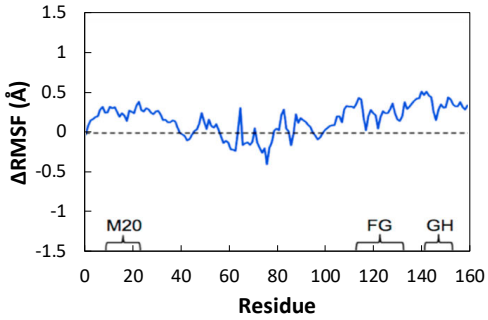


Figure 4.17: *Differences in the RMSFs between EcDHFR-N23PP/S148A and EcDHFR calculated for the backbone Cα atoms at the TSs.*

This result is in agreement with a recent evolutionary study of DHFRs from different species, which showed that mutations do not cause large changes to the equilibrium fluctuations. However, subtle changes to the equilibrium conformational sampling alter the free energy barrier of the enzymatic reaction.[98]

Deviations from classical Transition State Theory (TST), as a result of dynamic recrossings (κ) and quantum tunneling effects (γ), can be estimated by means of Ensemble-Averaged Variational Transition State Theory.[99-101] To correct the classical PMF and obtain the quasi-classical activation free energy (ΔG_{act}^{QC}), normal mode analyses were performed for the quantum region atoms. 13 TS structures were localized starting from different configurations of the corresponding simulation windows in the H and L variants of the wild type and mutant enzymes. After IRC calculations, we optimized the corresponding reactant structures and obtained the Hessian matrix for all the stationary structures. The final quantum mechanical vibrational corrections were obtained as an average over these structures resulting in statistically indistinguishable results

for the H and L variants. Once recrossing and tunneling transmission coefficients are computed we can incorporate this average into the exponential term as a correction to the effective classical free energy barrier giving a phenomenological free energy of activation, (ΔG_{eff}) that can be compared to the experimental one. Results are shown in Table 4.6.

Table 4.6: *Results from the QM/MM Simulations for Hydride Transfer in Light (L) and Heavy (H) EcDHFR-N23PP/S148A. Free energies are expressed in* $kcal \cdot mol^{-1}$*. Data for light and heavy wild type EcDHFR are from ref. 94. Experimental free energies are taken from the experimental results at pH 7.0 and 303 K.*

	κ	γ	ΔG_{eff}	ΔG_{exp}
L EcDHFR	0.57 ± 0.02	2.6 ± 0.5	14.35 ± 0.54	14.53 ± 0.01
H EcDHFR	0.49 ± 0.02		14.5 ± 0.5	14.58 ± 0.03
L EcDHFR-N23PP/S148A	0.53 ± 0.02	2.25 ± 0.45	16.6 ± 0.8	15.42 ± 0.02
H EcDHFR-N23PP/S148A	0.42 ± 0.02		16.7 ± 0.8	15.62 ± 0.02

To calculate the dynamic recrossing, we ran a 2 ns QM/MM-MD trajectory restrained to the transition state (TS) region saving one configuration every 10 ps. The resulting 200 configurations were used to compute unrestrained downhill trajectories. The velocity associated with the reaction coordinate is not properly thermalized in these initial configurations (because of the reaction coordinate restraint). Thus, following a procedure similar to that used by Gao and coworkers,[102] we selectively removed the projection of the velocity on the reaction coordinate and added a random value taken from a Maxwell-Boltzmann

distribution. For each of the saved TS configurations with modified velocities, we ran free MD simulations within the micro-canonical ensemble (NVE). Separately, for each configuration we integrated the equations of motion forwards and backwards, just changing the sign of the velocity components. Downhill trajectories were propagated from –2 ps to +2 ps using a time step of 0.05 fs. The trajectories obtained in the enzyme and in the solution were then classified as reactive trajectories when reactants connect to products (RP trajectories), or nonreactive trajectories leading either from reactants to reactants (RR) or from products to products (PP). To compute the transmission coefficient (κ), we used Equation 2.86 described in Chapter 2. The tunneling transmission coefficient (γ) was calculated with the small-curvature tunneling (SCT) approximation, which has been applied successfully to enzymatic reactions catalyzed by DHFR, yielding KIEs and phenomenological activation energies that are in good agreement with the experimental results.[103,104] The final tunneling contribution is obtained as the average over the reaction paths of 13 TS structures and is also statistically indistinguishable for H and L versions.

The calculations show clearly that the reduction in the rate constant in EcDHFR-N23PP/S148A relative to the wild type enzyme is essentially due to the increase in the ΔG_{act}^{QC} (see Table 4.6), while the recrossing transmission coefficients are similar but distinct and tunneling contributions are also very close in both wild type and variant enzyme. Comparison between H and L EcDHFR and EcDHFRN23PP/S148A shows that the difference in phenomenological rate constants between isotopomers arises from differences in the recrossing coefficients (κ).

Tunneling (γ) contributions are not affected by the change in mass between the L and H enzymes, which is not in agreement with proposals of tunneling enhancement by dynamic coupling.[105-107] According to our simulations, such compressive "promoting" motions do not need to be invoked to explain the observed changes in reactivity. However, the value of κ reflects the subtle coupling of protein environmental motions to the reaction coordinate in a way that is only apparent via a global description of all atomic positions. When the mass of the enzyme is increased, these motions are slower and the chemical system is not so efficient in relaxing to the reactant or product valleys after crossing the TS. This leads to an increase in the number of recrossings and therefore a reduction of the value of κ in the heavy enzyme compared to its light counterpart.

Simulations successfully reproduce the larger enzyme KIEs obtained experimentally for the variant. The increased enzyme KIE for EcDHFR-N23PP/S148A relative to the wild type, shown in Table 4.7, provides strong support for enhanced coupling of protein environmental motions to the reaction distinguished coordinate in the variant so that the chemical step becomes more sensitive to protein motions in the modified enzyme. These results show that the dynamic impact of protein motions on the chemical step is in fact larger in EcDHFR-N23PP/S148A than in the wild type enzyme.

Table 4.7: *Comparison between predicted and experimental hydride transfer rate constants and enzymatic KIES. Data for light and heavy wild type EcDHFR are from ref. 94.*

	Theory		**Experimental**	
	$k(s^{-1})$	k^L/k^H	$k(s^{-1})$	k^L/k^H
L EcDHFR	8.0	1.26 ± 0.04	47.2 ± 1.3	1.37 ± 0.03
H EcDHFR	6.3		34.4 ± 1.2	
L EcDHFR-N23PP/S148A	219	1.16 ± 0.04	209 ± 5	1.10 ± 0.04
H EcDHFR-N23PP/S148A	188		190 ± 8	

To gain insight into the role of protein motions, both the FESs for the hydride transfer catalyzed by EcDHFR, EcDHFR-N23PP/S148A and the FESs in the aqueous solution were obtained (see Figure 4.18) using the antisymmetric combination of distances and the antisymmetric combination of the electrostatic potential created by the environment on the donor and acceptor carbon atoms:

$$Ep = V_{C6}(\boldsymbol{r^N}) - V_{C4}(\boldsymbol{r^N}) \tag{4.13}$$

Umbrella sampling with force constants equal to $2500\, kJ \cdot mol^{-1} \cdot \text{Å}^{-2}$ and $0.024\, kJ^{-1} \cdot mol \cdot |e|^2$ were employed to obtain two-dimensional FESs. The FESs for the wild type and mutant enzymes were obtained from a total of 583 windows, while the FES for the water solution reaction needed a total of 1190 windows. Each window consisted in 5 ps of relaxation and 20 ps of production.

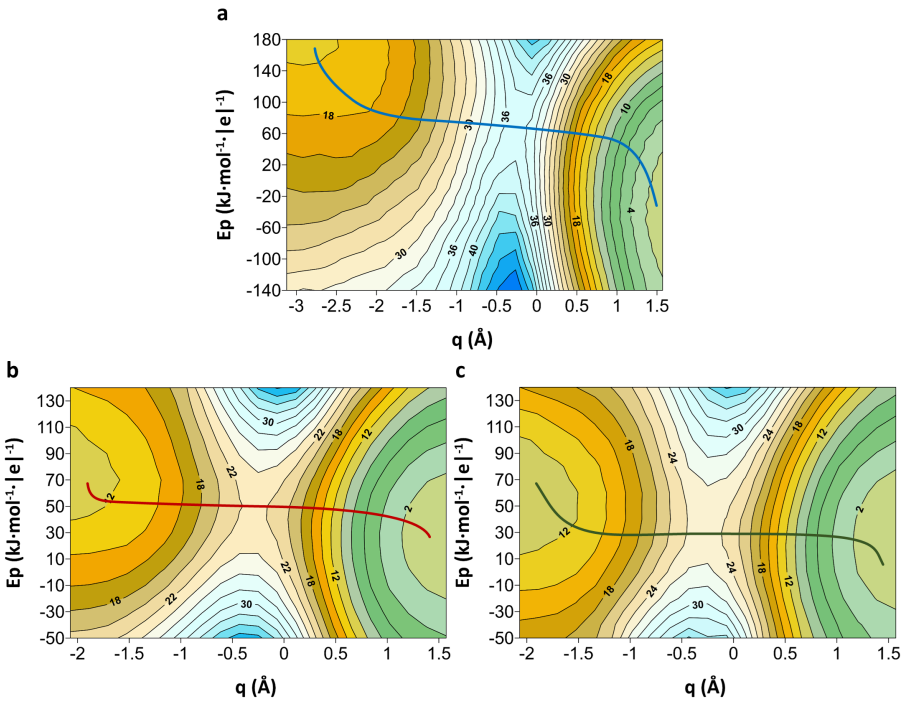


Figure 4.18: *Free energy surfaces corresponding to the aqueous solution reaction (a), the EcDHFR wild type (b) and the EcDHFR-N23PP/S148A mutant (c). Isocontour lines are drawn each 2 kcal · mol^{-1}. Continuous lines from reactants to products represent the minimum free energy path obtained on the free energy surfaces.*

A more complete description of the chemical process can be obtained by following the minimum free energy paths traced on the FESs (Figure 4.18). In all cases, the reaction starts with a change in the solvent coordinate, then essentially moves along the chemical coordinate to pass through the TS, and finishes with a new change in the solvent coordinate, leading to the relaxed products. Also, the behavior in the aqueous solution shows a larger change in the solvent coordinate, compared with the enzymatic systems. As in sections 4.1 and 4.2, environmental motions can

be characterized by means of the frequency associated with the solvent coordinate. To this end, the force constants and the effective masses were evaluated from the data obtained with the FESs in the same way than in the previous sections. The values of the force constants found for the enzymes are significantly larger than those for aqueous solution (see Table 4.8). This was expected, as is the same behavior found in the results obtained in sections 4.1 y 4.2 since deforming the environment is energetically more demanding in enzymes than in solution. As observed, the mass associated with the solvent coordinate is also significantly larger in the case of the two enzymes than in solution (see Table 4.8). The combined effect of the larger force constants and the larger associated mass for the enzymes means that the frequencies associated with the environmental motions are similar in all scenarios, and practically identical in EcDHFR and EcDHFR-N23PP/S148A (see Table 4.8). Thus, environmental motions relevant for the hydride transfer have similar time scales in the three cases (of the order of picoseconds or faster).

Table 4.8: *Force constants (K_s in $kcal^{-1} \cdot mol \cdot |e|^2$), effective masses ($m_s$ in $kcal^{-1} \cdot mol \cdot |e|^2 \cdot s^2$) and characteristic frequencies (v_s in cm^{-1}) associated with the solvent coordinate (s) for the hydride transfer reaction between solution, catalyzed by EcDHFR and by the mutant EcDHFR-N23PP/S148A.*

	Aqueous Sol.	**EcDHFR**	**EcDHFR-N23PP/S148A**
K_s	2.6×10^3	2.4×10^4	2.3×10^4
m_s	2.6×10^{30}	1.2×10^{31}	1.3×10^{31}
v_s	170	240	230

The characteristic frequencies obtained for the solvent coordinate in the variant DHFR are not consistent with a mutation-induced change of the protein dynamics that could have noticeable consequences for the rate of hydride transfer. This finding does however not exclude an effect of the mutations on the millisecond protein dynamics,[53] but such changes of the dynamics do not affect the barrier crossing during the hydride transfer.

4.4. Formate Dehydrogenase

NAD-dependent formate dehydrogenase (FDH) from the *methylotrophic bacterium Pseudomonas sp. 101* catalyzes the oxidation of a formate anion with the associated reduction of NAD^+ to NADH.

$$HCO_2^- + NAD^+ \rightarrow CO_2 + NADH \qquad (4.14)$$

The catalytic reaction involves cleavage of a carbon-hydrogen bond in the formate anion and the formation of a new one in the cofactor. [108]

FDH is a dimer with two chemically identical subunits with 391 residues each one and possessing independent active sites, as Figure 4.19 shows.[108]

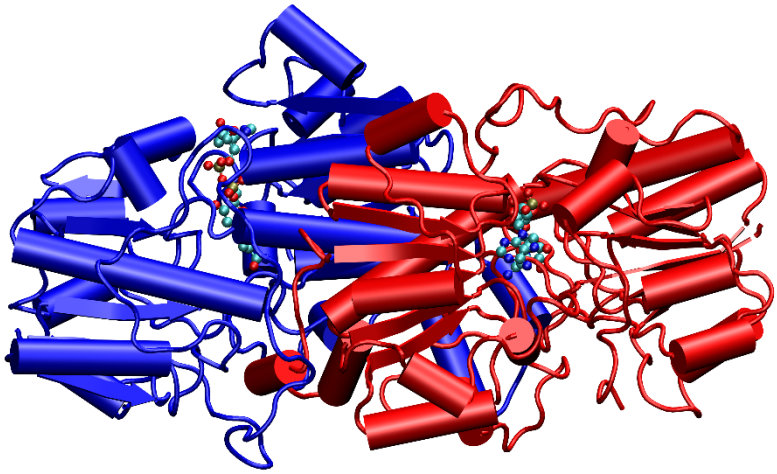


Figure 4.19: *Three-dimensional representation of FDH enzyme. Subunits are drawn in red and blue. Substrates in the active sites are depicted in balls and sticks format.*

The efficiency of this enzyme, its wide pH window and the irreversibility of the catalyzed reaction[109,110] make it an ideal catalyst to be used in chiral synthesis with NAD(P)$^+$-dependent oxidoreductases to regenerate the NADH cofactor.[84,111,112] Even so, its low operational stability and the high production cost of native FDHs limit their application with commercial purposes. Site-directed mutagenesis studies on FDH have been also carried out with the aim of transforming coenzyme specificity, as well as increasing its turnover frequency factor or its thermal stability.[113] Several computational approaches have been used to study the FDH catalytic mechanism at the molecular level,[114-118] aiming to provide new guidelines for future protein engineering.

The hydride transfer in the FDH catalyzed reaction seems to be the rate-limiting step.[119-121] The observed rate constant of 7.3 s^{-1}[122,123] indicates that the efficiency of the enzyme is relatively low compared with other NAD$^+$-dependent dehydrogenases, probably due to the strong interactions between the charged substrate and the polar environment of the active site.[117] Thus, it has been shown that the protein restrains the ground-state of the substrate in a conformation favorable to react, by means of a series of hydrogen bond interactions with the aminoacids of the active site.[114,124]

The temperature independence of the observed KIEs is assumed to be an indicator of a rigid active site and its optimization for tunneling.[124] Previous KIEs studies concluded that secondary hydrogen on the acceptor atom in the transition state is more loosely bonded than in either NAD$^+$ or NADH as the result of the coupling of the translational motion of

the hydrogen undergoing hydride transfer and the bending motion of the secondary hydrogen in the transition state.[120]

System Definition

In this case we took the coordinates and the system definition from previous studies done in our group.[114,115] To summarize, the initial coordinates of the protein were taken from the X-ray crystal structure of the Pseudomonas sp. complexed with the azide inhibitor in the formate binding site with PDB code 2NAD.[108] The azide molecule was replaced by a formate, hydrogen atoms were added to the full structure and the system was placed inside a water molecule cubic box of 80 Å, centered on the C4 atom of the cofactor in subunit B. Because of the size of the system, all the residues further than 38 Å from the C4 of the cofactor were removed (see Figure 4.20), and those further than 18 Å were kept frozen (20314 free atoms from a total of 23919 atoms).

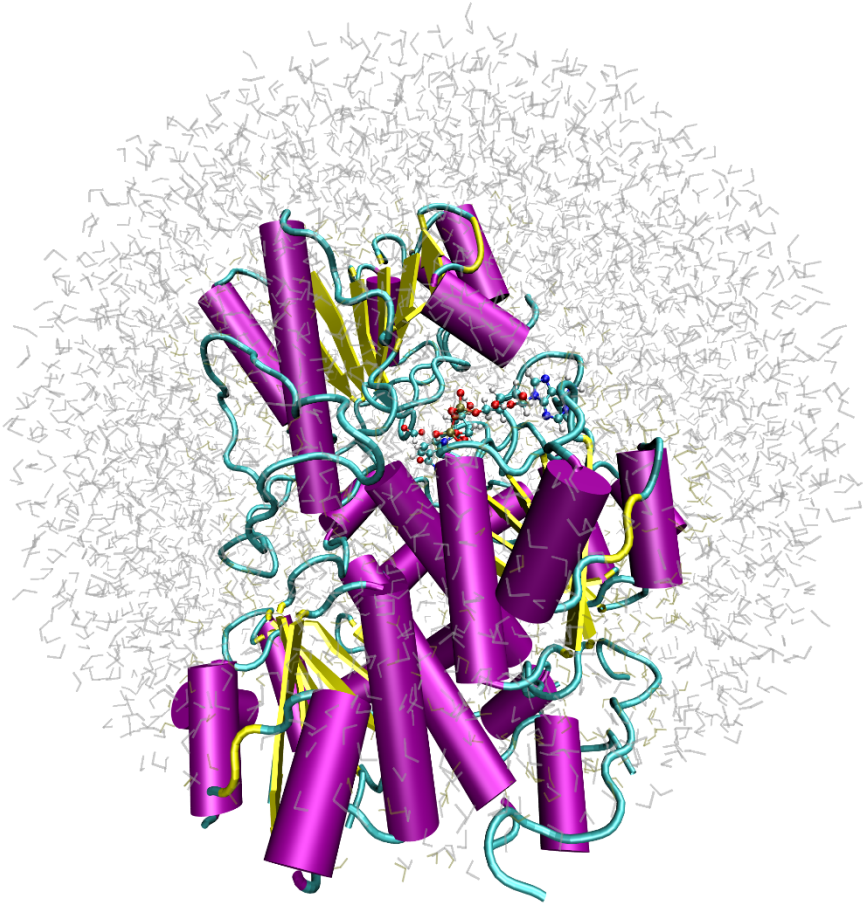


Figure 4.20: *Three-dimensional representation of FDH model used to perform the calculations. Formate and NAD^+ are depicted in balls and sticks format, α-helix in magenta, β-sheets in yellow and water molecules in lines format.*

The QM region consists of the formate anion together with the nicotinamide and ribose rings of the nicotinamide adenine dinucleotide of the cofactor subunit B (see Figure 4.21). These 33 atoms are described using the AM1 Hamiltonian.[37] The MM subsystem includes all the residues inside the 38 Å limit and all the water molecules. To treat the

nonbonding interactions, a switch function with a cutoff distance up to 12 Å was used.

In that study,[114] the RC was the antisymmetric combination of the distances of the hydrogen to the acceptor and to the donor ($q = d(CH41) - d(C4H41)$) and $q^{\ddagger} = -0.1$ Å is the value associated with the free energy maximum along this coordinate.[114] The system was then relaxed with 1000 ps of MD simulation at $q^{\ddagger}$.

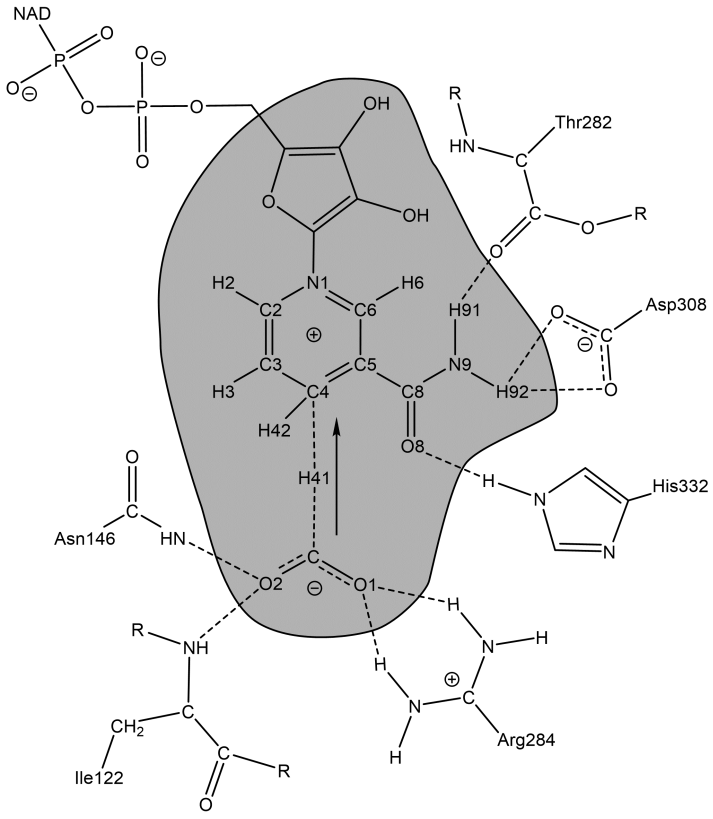


Figure 4.21: *Schematic representation of the FDH active site. The shaded part corresponds to the QM region. Atom labels are used as reference in the text.*

Results and Discussion

As a reaction that includes relevant motions from light particles, we have used the energy difference coordinate ($\Delta\varepsilon$) introduced in Chapter 3 as the RC:

$$\Delta\varepsilon = \epsilon(\boldsymbol{\zeta}, \boldsymbol{r_1}) - \epsilon(\boldsymbol{\zeta}, \boldsymbol{r_2}) \qquad (4.15)$$

where $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$ are the coordinates of the atoms selected to be quantized in the two states defined (see Figure 4.22), $\boldsymbol{\zeta}$ are the coordinates of the remaining atoms and $\epsilon(\boldsymbol{\zeta}, \boldsymbol{r})$ the total energy of the system.

The selected atoms used to define $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$ (see Figure 4.22)are the transferred hydride (H41) and the hydrogen bonded to the acceptor (H42) (see Figure 4.21 for the labels). To calculate the value of $\Delta\varepsilon$, for each of the configurations obtained from the MD calculations, the coordinates of these two hydrogen atoms were optimized, freezing the distance of H41 to the donor (C) or to the acceptor (C4) atoms at 1.1Å.

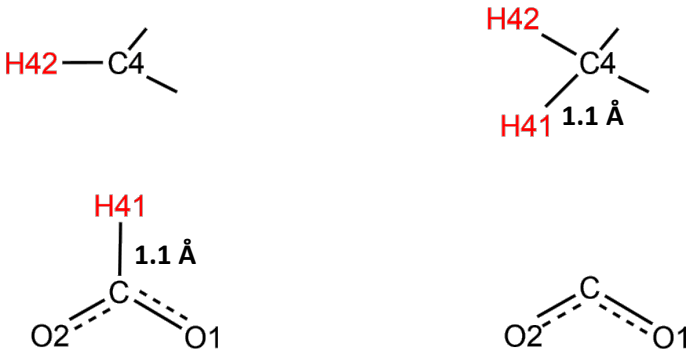


Figure 4.22: *Schematic representation of the* $\mathbf{r_1}$ *(left) and* $\mathbf{r_2}$ *(right) configurations. Selected atoms are shown in red.*

The system was equilibrated during 60 ps, with a time step of 0.5 fs, at $\Delta\varepsilon = 0\ kJ \cdot mol^{-1}$, using a force constant of $0.1\ kJ^{-1} \cdot mol$. For this purpose, a Langevin bath with a coupling temperature of 300 K was employed. A snapshot from this calculation was used as the initial configuration to perform simulations along the $\Delta\varepsilon$ coordinate. The corresponding PMF (see Figure 4.23) was obtained from a total of 46 windows with the $\Delta\varepsilon$ coordinate ranging from $-350\ kJ \cdot mol^{-1}$ to $100\ kJ \cdot mol^{-1}$ with a window width of $10\ kJ \cdot mol^{-1}$. Each window consisting in 5 ps of relaxation and 30 ps of production.

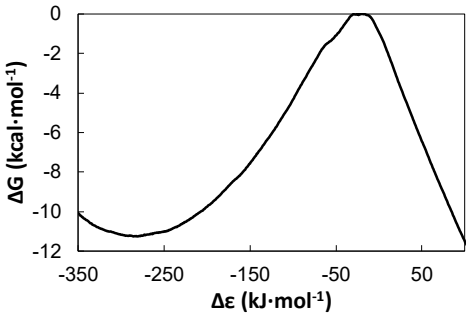


Figure 4.23: *Free energy profile along the Δε coordinate for the hydride transfer reaction catalyzed by FDH.*

The barrier obtained with $\Delta\varepsilon$ (11.2 $kcal \cdot mol^{-1}$) is close to that estimated in the previous study using the antisymmetric stretching of the hydride between the donor and the acceptor (q) as a reaction coordinate (12.4 $kcal \cdot mol^{-1}$).[114] The slightly smaller value obtained for the barrier is an indicator that for some configurations there is an energy barrier along the hydride transfer coordinate. This barrier is overcome when the hydride is placed in the middle with the antisymmetric stretching coordinate.

The adiabatic or non-adiabatic nature of the hydride transfer reaction is very sensitive to the donor acceptor distance ($Q = d(CC4)$).[125,126] Therefore, to explore its effect over the hydride transfer reaction a two dimensional PMF was obtained as a function of $\Delta\varepsilon$ and Q coordinates. Hybrid QM/MM molecular simulations were performed with the same characteristics as the previous ones. A total of 1380 simulations were carried out. For the $\Delta\varepsilon$ coordinate the same range of values and the same force constant than in the one dimensional PMF were employed. The Q coordinate was varied from 2.5 to 4.0 Å in 30 windows, using an

umbrella force constant of 2500 $kJ \cdot mol^{-1}$Å^{-2} for each particular value. For each simulation window 1 ps of relaxation Langevin MD followed by 10 ps of production were performed.

As observed in Figure 4.24, the reaction proceeds initially with a simultaneous change in $\Delta\varepsilon$ and Q coordinates, while at the TS the reaction proceeds along $\Delta\varepsilon$. The TS ensemble ($\Delta\varepsilon^{\ddagger} = 0\ kJ\ mol^{-1}$) displays a smooth dependence of the free energy on the Q coordinate, implying that the hydride transfer can take place at a significantly different donor acceptor distances.

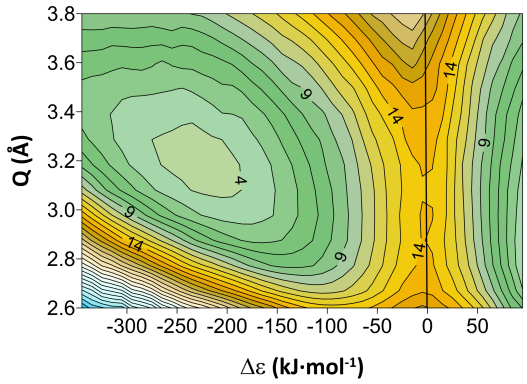


Figure 4.24: *Free energy surface of the hydride transfer reaction catalyzed by FDH. Isocontour lines are drawn each 1 $kcal \cdot mol^{-1}$. Black line represents the TS ensemble.*

As discussed before, the hydride transfer reaction nature is strongly dependent on Q. Thus, assuming a classical behavior for the Q coordinate,[126] the rate constant can be obtained as a weighted average of the rate constant (adiabatic or non-adiabatic) associated to each value of Q:

$$< k_r >= \int_0^{\infty} \rho(Q; \Delta\varepsilon^{\ddagger}) k_r(Q) dQ \qquad (4.16)$$

From the two dimensional PMF (Figure 4.24), we can obtain the free energy profile along the Q coordinate at $\Delta\varepsilon^{\ddagger}$ ($\Delta G\left(Q; \Delta\varepsilon^{\ddagger}\right)$). The probability distribution can then be calculated from those data as follows:

$$\rho(Q; \Delta\varepsilon^{\ddagger}) = \frac{e^{-\frac{\Delta G\left(Q; \Delta\varepsilon^{\ddagger}\right)}{RT}}}{\int_0^{\infty} e^{-\frac{\Delta G\left(Q; \Delta\varepsilon^{\ddagger}\right)}{RT}} dQ} \qquad (4.17)$$

At this point, it can be very illustrative to compare the nature of the TS ensembles obtained using as RC the $\Delta\varepsilon$ or q coordinates. With this purpose, we performed a 100 ps MD simulation with an umbrella constraint at $q^{\ddagger}$ obtaining the corresponding distribution of donor-acceptor distances at the TS $\rho\left(Q; q^{\ddagger}\right)$. The results show that restraining the hydrogen to be in between the two carbons limits the range of values that the Q coordinate can reach (see Figure 4.25) and reduces the average distance at the TS (2.66 ± 0.06 Å [114] versus 3.1 ± 0.1 Å with the $\Delta\varepsilon$ coordinate), thus changing the donor-acceptor distance at which the transfer can happen.

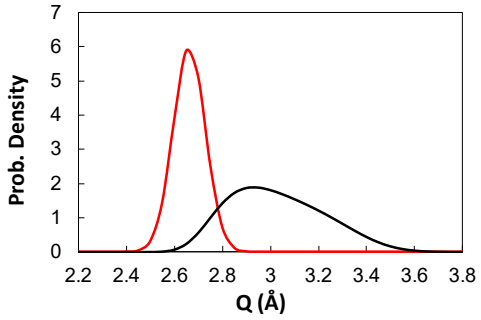


Figure 4.25: *Probability distribution of the Q coordinate at the TS using $\Delta\varepsilon$ (black line) and q (red line) to define the TS ensemble.*

On the other hand the rate constant $k_r(Q)$ can be expressed as the average one-way flux in the solvent coordinate through the TS as described in Chapter 2.

$$k_r(Q) = g(C) \cdot e^{-\frac{\Delta G^{\ddagger}_{ZPE}(Q)}{RT}} \tag{4.18}$$

$$g(C) = \frac{\sqrt{k_s \cdot m_s}}{2\pi \cdot RT} \int d\dot{S} \cdot \kappa_{LZ}(C) \cdot \dot{S} \cdot \theta(\dot{S}) \cdot e^{-\frac{m_S \dot{S}^2}{2 \cdot RT}} \tag{4.19}$$

To calculate the ZPE correction and the coupling at the TS, we took snapshots at Q coordinate values from 2.6 to 3.3 Å (with an interval of 0.1 Å), using the trajectories computed at $\Delta\varepsilon^{\ddagger}$. These snapshots were used as starting coordinates to perform instantaneous potential energy surfaces over the hydrogen transfer coordinate q and the bending motion of the secondary hydrogen (see Figure 4.26), keeping frozen the remaining

degrees of freedom. In addition, in order to compute the ZPE at the RS we selected snapshots from the corresponding trajectory.

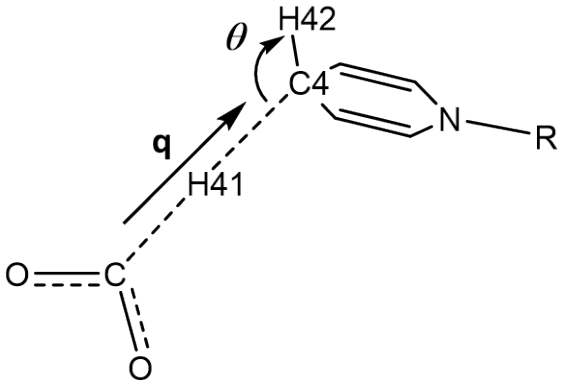


Figure 4.26: *Representation of the coordinates used to calculate the potential energy surfaces for the hydrogen transfer at* $\Delta\varepsilon^{\ddagger}$.

Once calculated, these surfaces were averaged for each value of Q (see Figure 4.27a). Then, they were used to compute the vibrational levels associated to these two coordinates by means of the DVR methodology[127,128] (see Figure 4.27b). From these calculations we took the ZPE at different Q values to include them in the calculation of the rate constant. The same calculation was performed over the surfaces from RS snapshots to obtain the ZPE correction to the free energy barrier ($\Delta ZPE = ZPE_{TS} - ZPE_{RS}$).

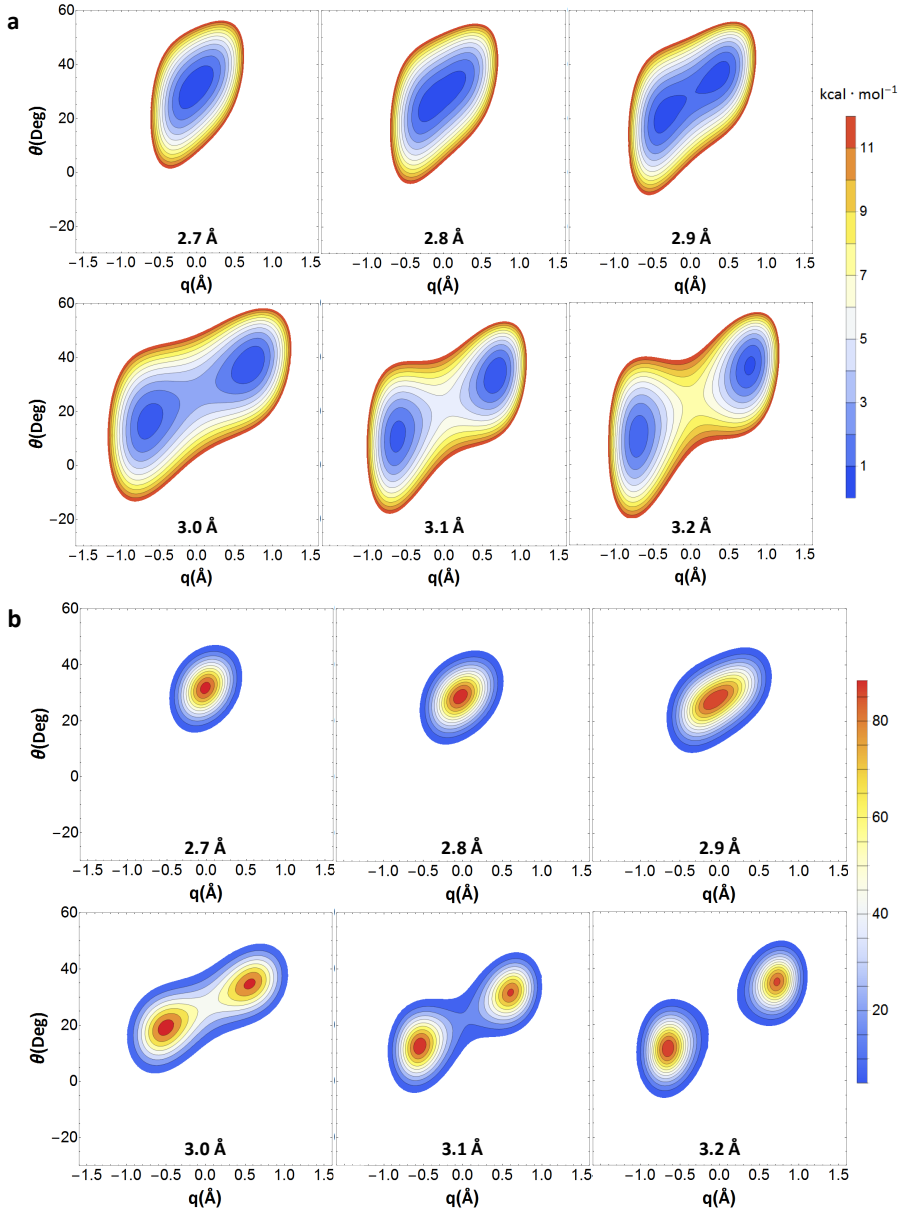


Figure 4.27: *Averaged PESs at Q values of 2.7, 2.8, 2.9, 3.0, 3.1 and 3.2 Å (from left to right) as a function of the bending motion of the secondary hydrogen and the q coordinate* (a). *First vibrational wavefunction probability distribution calculated from the PESs* (b).

The analysis of the averaged PESs shows that the hydride transfer takes place in all cases with a coupled motion along the stretching and bending coordinates. The motion of the secondary hydrogen atom is thus coupled to the primary one. In addition, we can observe a change in the nature of the hydride transfer as the donor-acceptor distance increases. At short distances, the PES shows a minimum when the hydride is placed in between these two atoms. As the Q distance increases, a barrier appears for the hydride transfer. If the ground vibrational level lies below the barrier, then the hydride transfer will take place through a potential energy barrier, via non-adiabatic tunneling. Figure 4.28 shows that at donor-acceptor distances smaller than 3.0 Å tunneling cannot happen because the ZPE level is above the barrier.

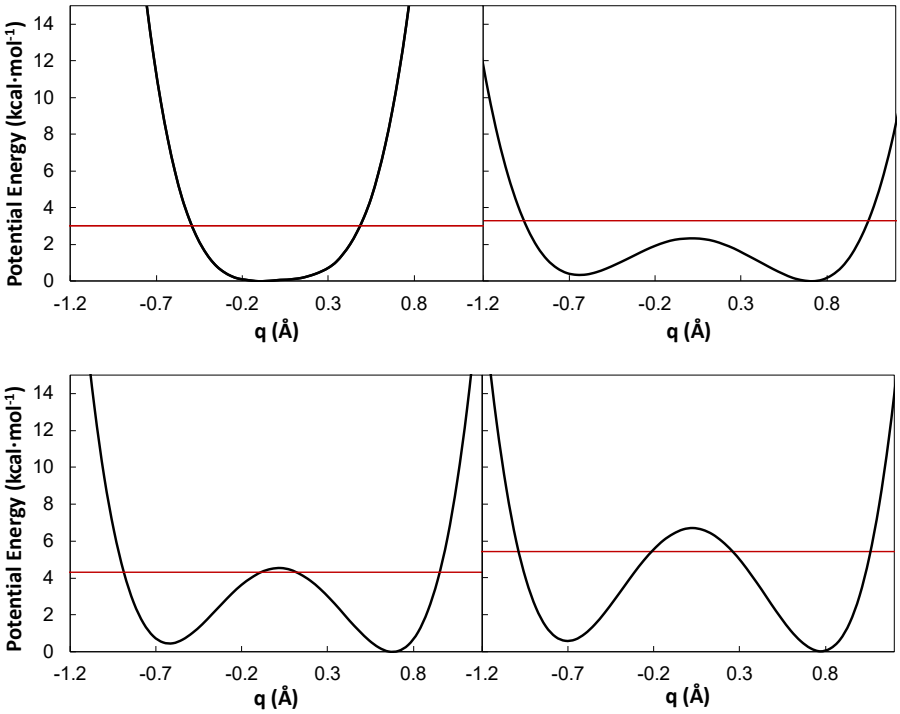


Figure 4.28: *Potential energy profiles at Q values of 2.8, 3.0, 3.1 and 3.2 Å. Red line represents the ZPE value. The profile is taken from a slice of the previous PESs (see Figure 4.27) taking the minimum energy point at each value of q.*

The coupling appearing in equation (4.19) is calculated from the splitting between the two first vibrational eigenvalues (Λ).[126] At large values of Q, the diabatic states localized in each well are weakly coupled and C tends to zero, while for intermediate Q values, the barrier will be lower and thinner, resulting in a larger splitting of the eigenlevels and causing an enhancement of tunneling. As long as the two first vibrational eigenvalues are well separated from the next upper levels, the coupling can be calculated as:

$$C = \frac{\Lambda}{2} \tag{4.20}$$

Finally, at small values of Q, the proton barrier can be low enough so that the ground vibrational eigenstates is found above the barrier, or there is no barrier at all, which corresponds to the adiabatic limit. The dependence of the coupling with Q was approximated by an exponential function[126] as shown in Figure 4.29a.

In Figure 4.29a the coupling is close to zero for Q distances larger than 3.0 Å. This indicates that hydride transfer will not happen at those distances and therefore tunneling contribution to the observed rate constant is expected to be small. Results displayed in Figure 4.29b show that at short values of Q the contribution to the barrier ΔZPE is small since the PES along the transfer coordinate at the TS, for those distances, is a single well similar to the PES on the RS. Thus the contribution of ZPE to the barrier is small. As the Q value increases the PES at the TS becomes wider and the ZPE_{TS} decreases. After 2.9 Å, the PES splits into two narrower wells and the ZPE_{TS} increases again. When the donor-acceptor distance becomes larger the ZPE_{TS} becomes nearly constant and the contribution to the barrier does not change significantly with Q.

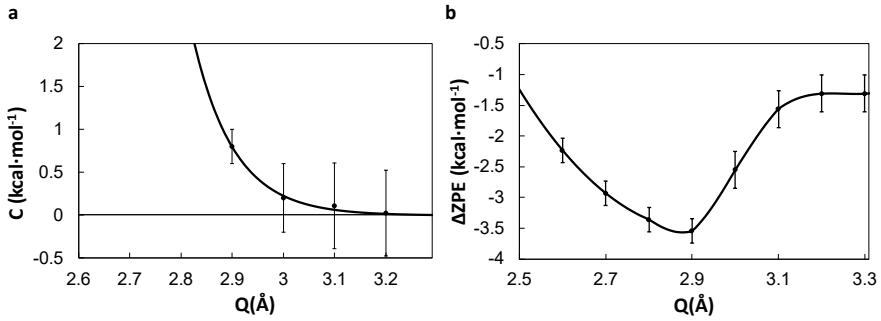


Figure 4.29: *Dependence of the ZPE contribution to the barrier with Q (*a*) and of the coupling between the localized vibrational states of the hydrogen atoms (*b*).*

Once the probability distribution of donor acceptor distances at the TS, the ZPE correction to the free energy barrier and the coupling were calculated; the rate constant along the Q coordinate (see equation (4.18)) and the averaged rate constant (see equation (4.16)) can be estimated. For comparison, the rate constant was also calculated for the adiabatic limit, in which no tunneling is considered, using the following equation explained in Chapter 2:

$$k_{r,ad}(Q) = \frac{\omega_s}{2\pi} \cdot e^{-\frac{\Delta G^{\ddagger}_{ZPE}(Q)}{RT}} \qquad (4.21)$$

Figure 4.30 shows the value of the rate constant obtained as a function of the value of the donor-acceptor distance at which the hydride transfer takes place for the complete treatment (equation (4.18)) and for the adiabatic limit (equation (4.21)). Each rate constant profile is weighted by the donor-acceptor distance probability distribution at the TS as determined using the q coordinate for the adiabatic limit ($\rho(Q; q^{\ddagger})$) and the $\Delta\varepsilon$ coordinate for the complete treatment ($\rho(Q; \Delta\varepsilon^{\ddagger})$). There is a

clear displacement in the maximum of each profile mainly due to the difference between the different probability distribution of the donor-acceptor distance. This suggest that, for each treatment, the largest contribution to the total rate constant comes from events taken place at different distances (~2.7 Å for the adiabatic limit and ~2.8 Å for the complete description).

In the case of the full treatment, where the system can reach larger Q distances, the contribution from distances $Q > 3.0$ Å, where tunneling can be the dominant regime for hydride transfer, is very small. As said before this is mainly due to the small coupling displayed by the system at those distances.

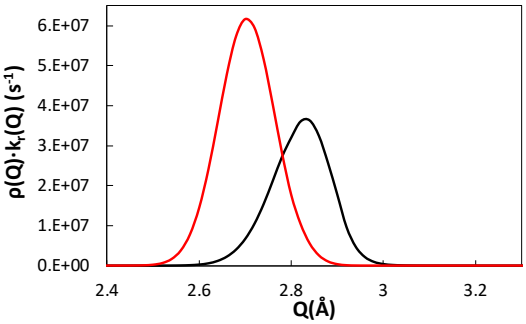


Figure 4.30: *Profiles along the Q coordinate of the rate constant weighted with the donor acceptor distance probability distribution for the adiabatic limit (red line) and the full treatment (black).*

Once the formalism to obtain the rate constant has been presented, we can use it to calculate KIEs and to compare the predicted values to the experimental ones.[120,124] KIEs are very sensitive to the TS description while they do not depend on the classical free energy difference between the TS and the reactants, a magnitude affected by systematic errors due

the use of semiempirical methods. In our treatment KIEs were obtained as:

$$KIE = \frac{k_{r,X'Y'}}{k_{r,XY}} = \frac{\int_0^\infty \rho(Q;\Delta\varepsilon^\ddagger)k_{r,X'Y'}(Q)dQ}{\int_0^\infty \rho(Q;\Delta\varepsilon^\ddagger)k_{r,XY}(Q)dQ} \qquad (4.22)$$

The isotopic substitution affects ZPEs and the coupling, which were recalculated with different masses for the two quantized hydrogen atoms: HH, DH, HD, DD and TH. The notation is: H for hydrogen, D for deuterium and T for tritium. The first letter refers to the transferred hydrogen and the second letter to the secondary hydrogen.

The coupling profiles displayed in Figure 4.31a show that the heavier isotopes present the lower coupling values. This is expected, since the nuclear wavefunction overlap decreases when increasing the mass of the atom. On the other hand, from the ZPE energy profiles shown in Figure 4.31b it can be seen that as the mass of the primary hydrogen atom increases the magnitude for the ZPE correction to the classical barrier decreases, resulting in a normal isotope effect. However, when the mass of the secondary hydrogen atom is increased, the behavior is significantly different: at short Q distances the magnitude of the ZPE correction is larger for HD than for HH indicating and inverse isotope effect. This is due to the fact that at small donor-acceptor distances the force constant for the bending motion increases at the TS due to the vicinity between the two fragments. This effects disappears as Q increases and ZPE corrections become very similar for HD and HH at large Q distances.

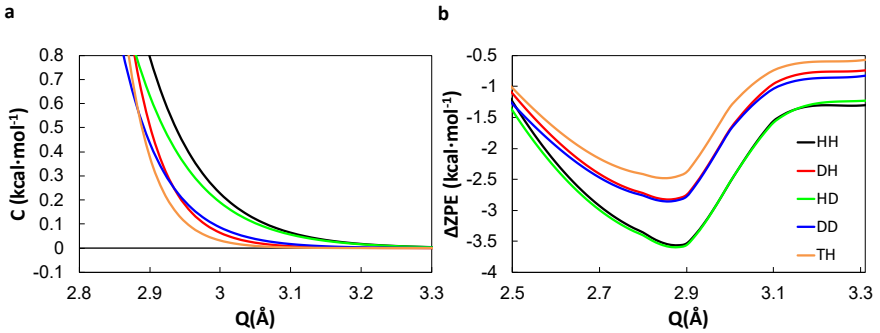


Figure 4.31: *Profiles of the ZPE contribution to the barrier (*a*) and the coupling (*b*) for different isotopic substitutions.*

The rate constant as a function of the Q distance in our treatment shows smaller values for the heavy isotopic substitutions of the primary hydrogen (see Figure 4.32). However for the HD combination the profile shows a slight increase in the rate constant at short Q distance in comparison with the HH counterpart, influenced by the ZPE behavior along the Q coordinate commented above. We can observe that although the system can access Q distances larger than 3.0 Å at the TS, none of the profiles has a significantly higher rate constant at those distances. Therefore an increase in the Q distances accessible to the system, due for example an increase in temperature, should not change the relation between the different profiles and thus the KIEs should be, in principle, temperature independent, as observed experimentally.[124]

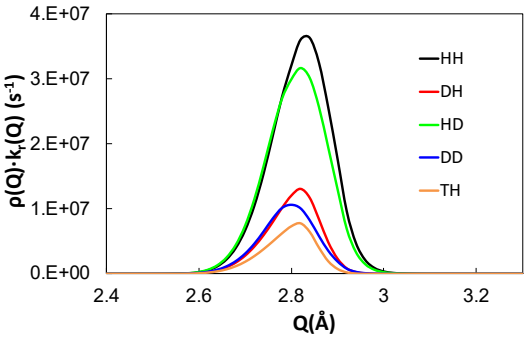


Figure 4.32: *Profiles along the Q coordinate of the weighted rate constant for different isotopic substitutions.*

The calculated KIEs were compared to those measured experimentally by Cleland et al.[120] and Kohen et al.[124] (see Table 4.9). A comparison of our calculated primary and secondary KIEs with the experimental data reported in the literature reveals a very good agreement, except for the secondary KIE which is slightly underestimated in our treatment. The large secondary KIE, observed in this system, has been used as an indicator of the coupled motion between the primary and secondary hydrogen atoms, but also as a sign that tunneling can be important for the hydride transfer in FDH.[120] This might suggest an underestimation of the tunneling contribution to the rate constant in our model.

Table 4.9: *Kinetic isotope effects theoretically and experimentally calculated.*

KIE	Theoretical Values	Experimental values	
		Ref. 120	**Ref. 124**
HH/DH	3.3 ± 0.3	2.3 ± 0.2	3.41 ± 0.3
HH/HD	1.12 ± 0.04	1.23 ± 0.03	–
HH/DD	3.6 ± 0.4	–	–
HH/TH	5.7 ± 0.7	–	5.93 ± 0.73
DH/DD	1.13 ± 0.05	1.07 ± 0.02	–
DH/TH	1.8 ± 0.6	–	1.70 ± 0.06

In order to understand the influence of the non-adiabatic regime to the KIEs we repeated our calculations in the adiabatic limit, this is using equation (4.21). The results shown in Table 4.10 indicate that the inclusion of the non-adiabatic hydride transfer, does not yield significant differences in the calculation of primary KIEs. However, the secondary KIE is sensitive to the inclusion of the non-adiabatic regime for the hydride transfer improving the agreement with experimental data. This would demonstrate that the consideration of tunneling for the coupled motions of the primary and secondary hydrogen atoms during the transfer can be important to explain some experimental observations in FDH.

Table 4.10: *Kinetic isotope effects in the complete treatment and in the adiabatic limit.*

KIE	Complete treatment	Adiabatic limit
HH/DH	3.3 ± 0.3	3.3 ± 0.3
HH/HD	1.12 ± 0.04	0.96 ± 0.04
HH/DD	3.4 ± 0.4	3.0 ± 0.4
HH/TH	5.9 ± 0.7	5.7 ± 0.7
DH/DD	1.13 ± 0.05	0.94 ± 0.05
DH/TH	1.8 ± 0.6	1.7 ± 0.6

Bibliography

(1) Field, J. A.; Sierra-Alvarez, R. *Rev. Environ. Sci. Biotechnol.* **2004**, *3*, 185.

(2) Hughes, K.; Meek, M. E.; Caldwell, I. *J. Environ. Sci. Heal. C* **1994**, *12*, 293.

(3) Stucki, G.; Thueer, M. *Environ. Sci. Technol.* **1995**, *29*, 2339.

(4) Janssen, D. B.; Scheper, A.; Dijkhuizen, L.; Witholt, B. *Appl. Environ. Microbiol.* **1985**, *49*, 673.

(5) Verschueren, K. H.; Seljée, F.; Rozeboom, H. J.; Kalk, K. H.; Dijkstra, B. W. *Nature* **1993**, *363*, 693.

(6) Ridder, I. S.; Rozeboom, H. J.; Dijkstra, B. W. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **1999**, *55*, 1273.

(7) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins* **2005**, *61*, 704.

(8) Bas, D. C.; Rogers, D. M.; Jensen, J. H. *Proteins* **2008**, *73*, 765.

(9) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. *J. Chem. Theory Comput.* **2011**, *7*, 525.

(10) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. *J. Chem. Theory Comput.* **2011**, *7*, 2284.

(11) Field, M. J. *A Practical Introduction to the Simulation of Molecular Systems*; Cambridge University Press, **1999**.

(12) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209.

(13) Devi-Kesavan, L. S.; Gao, J. *J. Am. Chem. Soc.* **2003**, *125*, 1532.

(14) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(15) Neria, E.; Fischer, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 1902.

(16) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.

(17) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.

(18) Jorgensen, W. L.; Ulmschneider, J. P.; Tirado-Rives, J. *J. Phys. Chem. B* **2004**, *108*, 16264.

(19) Gao, J.; Xia, X. *J. Am. Chem. Soc.* **1993**, *115*, 9667.

(20) Ruiz-Pernía, J. J.; Martí, S.; Moliner, V.; Tuñón, I. *J. Chem. Theory Comput.* **2012**, *8*, 1532.

(21) Schanstra, J. P.; Kingma, J.; Janssen, D. B. *J. Biol. Chem.* **1996**, *271*, 14747.

(22) Shurki, A.; Štrajbl, M.; Villà, J.; Warshel, A. *J. Am. Chem. Soc.* **2002**, *124*, 4097.

(23) Soriano, A.; Silla, E.; Tuñón, I.; Martí, S., S.; Moliner, V.; Bertrán, J. *Theor. Chem. Accounts* **2004**, *112*, 327.

(24) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Accounts* **2008**, *120*, 215.

(25) Hwang, J. K.; King, G.; Creighton, S.; Warshel, A. *J. Am. Chem. Soc.* **1988**, *110*, 5297.

(26) Soriano, A.; Silla, E.; Tuñón, I.; Ruiz-López, M. F. *J. Am. Chem. Soc.* **2005**, *127*, 1946.

(27) Gertner, B. J.; Wilson, K. R.; Hynes, J. T. *J. Chem. Phys.* **1989**, *90*, 3537.

(28) Gertner, B. J.; Bergsma, J. P.; Wilson, K. R.; Lee, S.; Hynes, J. T. *J. Chem. Phys.* **1987**, *86*, 1377.

(29) Bergsma, J. P.; Gertner, B. J.; Wilson, K. R.; Hynes, J. T. *J. Chem. Phys.* **1987**, *86*, 1356.

(30) Vidgren, J.; Svensson, L. A.; Liljas, A. *Nature* **1994**, *368*, 354.

(31) Schlissel, M. S.; Brown, D. D. *Cell* **1984**, *37*, 903.

(32) Thoma, F.; Koller, T.; Klug, A. *J. Cell. Biol.* **1979**, *83*, 403.

(33) Ruottinen, H. M.; Rinne, U. K. *J. Neurol.* **1998**, *245*, 25.

(34) Goetz, C. G. *Neurology* **1998**, *50*, S26.

(35) Vidgren, J.; Ovaska, M.; Tenhunen, J.; Tilgmann, C.; Lotta, T.; Männistö, P. T. In *S-Adenosylmethionine-Dependent Methyltransferases: Structures And Functions*; Cheng, X., Blumenthal, R. M., Eds.; World Scientific Publishing Co. Pte. Ltd.: Singapore, **1999**, p 55.

(36) Zheng, Y.; Bruice, T. C. *J. Am. Chem. Soc.* **1997**, *119*, 8137.

(37) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

(38) Roca, M.; Moliner, V.; Ruiz-Pernía, J. J.; Silla, E.; Tuñón, I. *J. Phys. Chem. B* **2005**, *110*, 503.

(39) Ruiz-Pernía, J. J.; Silla, E.; Tuñón, I.; Martí, S. *J. Phys. Chem. B* **2006**, *110*, 17663.

(40) Kästner, J. *J. Chem. Phys.* **2009**, *131*, 034109.

(41) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. *Chem. Rev.* **2006**, *106*, 3210.

(42) Warshel, A. *J. Biol. Chem.* **1998**, *273*, 27035.

(43) Roca, M.; Martí, S.; Andrés, J.; Moliner, V.; Tuñón, I.; Bertrán, J.; Williams, I. H. *J. Am. Chem. Soc.* **2003**, *125*, 7726.

(44) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.

(45) Rodgers, J.; Femec, D. A.; Schowen, R. L. *J. Am. Chem. Soc.* **1982**, *104*, 3263.

(46) Ruggiero, G. D.; Williams, I. H.; Roca, M.; Moliner, V.; Tuñón, I. *J. Am. Chem. Soc.* **2004**, *126*, 8634.

(47) Kanaan, N.; Ruiz-Pernía, J. J.; Williams, I. H. *Chem. Commun.* **2008**, 6114.

(48) Zhang, J.; Klinman, J. P. *J. Am. Chem. Soc.* **2011**, *133*, 17134.

(49) Hegazi, M. F.; Borchardt, R. T.; Schowen, R. L. *J. Am. Chem. Soc.* **1979**, *101*, 4359.

(50) Gray, C. H.; Coward, J. K.; Schowen, K. B.; Schowen, R. L. *J. Am. Chem. Soc.* **1979**, *101*, 4351.

(51) Roca, M.; Moliner, V.; Tuñón, I.; Hynes, J. T. *J. Am. Chem. Soc.* **2006**, *128*, 6186.

(52) Roca, M.; Andrés, J.; Moliner, V.; Tuñón, I.; Bertrán, J. *J. Am. Chem. Soc.* **2005**, *127*, 10648.

(53) Bhabha, G.; Lee, J.; Ekiert, D. C.; Gam, J.; Wilson, I. A.; Dyson, H. J.; Benkovic, S. J.; Wright, P. E. *Science* **2011**, *332*, 234.

(54) Mullen, R. G.; Shea, J. E.; Peters, B. *J. Chem. Phys.* **2014**, *140*, 041104.

(55) Blakley, R. L.; Benkovic, S. J. *Chemistry and biochemistry of pterins*; J. Wiley: New York, **1985**.

(56) Blakley, R. L.; Benkovic, S. J. *Chemistry and biochemistry of folates*; Wiley: New York, **1984**.

(57) Sawaya, M. R.; Kraut, J. *Biochemistry* **1997**, *36*, 586.

(58) Sandler, M.; Smith, H. J. *Design of enzyme inhibitors as drugs*; Oxford University Press: Oxford England ; New York, **1989**.

(59) Calvert, A. H.; Jones, T. R.; Dady, P. J.; Grzelakowska-Sztabert, B.; Paine, R. M.; Taylor, G. A.; Harrap, K. R. *Eur. J. Cancer* **1980**, *16*, 713.

(60) Miller, G. P.; Benkovic, S. J. *Biochemistry* **1998**, *37*, 6327.

(61) Cameron, C. E.; Benkovic, S. J. *Biochemistry* **1997**, *36*, 15792.

(62) Ruiz-Pernía, J. J.; Luk, L. Y. P.; García-Meseguer, R.; Martí, S.; Loveridge, E. J.; Tuñón, I.; Moliner, V.; Allemann, R. K. *J. Am. Chem. Soc.* **2013**, *135*, 18689.

(63) Adamczyk, A. J.; Cao, J.; Kamerlin, S. C. L.; Warshel, A. *Proc. Nat. Acad. Sci. USA* **2011**, *108*, 14115.

(64) Miller, G. P.; Benkovic, S. J. *Biochemistry* **1998**, *37*, 6336.

(65) Boehr, D. D.; McElheny, D.; Dyson, H. J.; Wright, P. E. *Science* **2006**, *313*, 1638.

(66) Osborne, M. J.; Schnell, J.; Benkovic, S. J.; Dyson, H. J.; Wright, P. E. *Biochemistry* **2001**, *40*, 9846.

(67) Arora, K.; Brooks III, C. L. *J. Am. Chem. Soc.* **2009**, *131*, 5642.

(68) Loveridge, E. J.; Behiry, E. M.; Guo, J.; Allemann, R. K. *Nat. Chem.* **2012**, *4*, 292.

(69) Bolin, J. T.; Filman, D. J.; Matthews, D. A.; Hamlin, R. C.; Kraut, J. *J. Biol. Chem.* **1982**, *257*, 13650.

(70) Benkovic, S. J.; Hammes-Schiffer, S. *Science* **2003**, *301*, 1196.

(71) Appleman, J. R.; Howell, E. E.; Kraut, J.; Kühl, M.; Blakley, R. L. *J. Biol. Chem.* **1988**, *263*, 9187.

(72) Appleman, J. R.; Howell, E. E.; Kraut, J.; Blakley, R. L. *J. Biol. Chem.* **1990**, *265*, 5579.

(73) Rajagopalan, P. T. R.; Lutz, S.; Benkovic, S. J. *Biochemistry* **2002**, *41*, 12618.

(74) Karginov, V. A.; Mamaev, S. V.; An, H.; Van Cleve, M. D.; Hecht, S. M.; Komatsoulis, G. A.; Abelson, J. N. *J. Am. Chem. Soc.* **1997**, *119*, 8166.

(75) Dion, A.; Linn, C. E.; Bradrick, T. D.; Georghiou, S.; Howell, E. E. *Biochemistry* **1993**, *32*, 3479.

(76) Howell, E. E.; Booth, C.; Farnum, M.; Kraut, J.; Warren, M. S. *Biochemistry* **1990**, *29*, 8561.

(77) Villafranca, J. E.; Howell, E. E.; Voet, D. H.; Strobel, M. S.; Ogden, R. C.; Abelson, J. N.; Kraut, J. *Science* **1983**, *222*, 782.

(78) Howell, E. E.; Villafranca, J. E.; Warren, M. S.; Oatley, S. J.; Kraut, J. *Science* **1986**, *231*, 1123.

(79) Dunn, S. M. J.; Lanigan, T. M.; Howell, E. E. *Biochemistry* **1990**, *29*, 8569.

(80) Pu, J.; Ma, S.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. B* **2005**, *109*, 8551.

(81) Fierke, C. A.; Johnson, K. A.; Benkovic, S. J. *Biochemistry* **1987**, *26*, 4085.

(82) Pu, J.; Ma, S.; García-Viloca, M.; Gao, J.; Truhlar, D. G.; Kohen, A. *J. Am. Chem. Soc.* **2005**, *127*, 14879.

(83) Thorpe, I. F.; Brooks, C. L. *J. Phys. Chem. B* **2003**, *107*, 14042.

(84) Tishkov, V. I.; Popov, V. O. *Biomol. Eng.* **2006**, *23*, 89.

(85) Liu, C. T.; Layfield, J. P.; Stewart, R. J.; French, J. B.; Hanoian, P.; Asbury, J. B.; Hammes-Schiffer, S.; Benkovic, S. J. *J. Am. Chem. Soc.* **2014**, *136*, 10349.

(86) Li, L.; Wright, P. E.; Benkovic, S. J.; Falzone, C. J. *Biochemistry* **1992**, *31*, 7826.

(87) Huennekens, F. M.; Scrimgeour, K. G. In *Pteridine chemistry : proceedings of the third international symposium held at the Institut für Organische Chemie der Technischen Hochschule Stuttgart, September 1962*; Pergamon Press: Oxford, **1964**.

(88) Khavrutskii, I. V.; Price, D. J.; Lee, J.; Brooks, C. L. *Protein Sci.* **2007**, *16*, 1087.

(89) Liu, H.; Warshel, A. *J. Phys. Chem. B* **2007**, *111*, 7852.

(90) Nashine, V. C.; Hammes-Schiffer, S.; Benkovic, S. J. *Curr. Opin. Chem. Biol.* **2010**, *14*, 644.

(91) Wang, L.; Tharp, S.; Selzer, T.; Benkovic, S. J.; Kohen, A. *Biochemistry* **2006**, *45*, 1383.

(92) Allemann, R. K.; Evans, R. M.; Loveridge, E. J. *Biochem. Soc. Trans.* **2009**, *37*, 349.

(93) Sikorski, R. S.; Wang, L.; Markham, K. A.; Rajagopalan, P. T. R.; Benkovic, S. J.; Kohen, A. *J. Am. Chem. Soc.* **2004**, *126*, 4778.

(94) Luk, L. Y. P.*et al. Proc. Nat. Acad. Sci. USA* **2013**, *110*, 16344.

(95) Frisch, M. J.*et al.*; Gaussian, Inc.: Wallingford, CT, USA, **2009**.

(96) Vardi-Kilshtain, A.; Major, D. T.; Kohen, A.; Engel, H.; Doron, D. *J. Chem. Theo. Com.* **2012**, *8*, 4786.

(97) García-Viloca, M.; Truhlar, D. G.; Gao, J. *Biochemistry* **2003**, *42*, 13558.

(98) Liu, C. T.; Hanoian, P.; French, J. B.; Pringle, T. H.; Hammes-Schiffer, S.; Benkovic, S. J. *Proc. Nat. Acad. Sci. USA* **2013**, *110*, 10159.

(99) Truhlar, D. G.; Gao, J.; Alhambra, C.; García-Viloca, M.; Corchado, J.; Sánchez, M. L.; Villà, J. *Acc. Chem. Res.* **2002**, *35*, 341.

(100) Alhambra, C.; Corchado, J.; Sánchez, M. L.; García-Viloca, M.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. B* **2001**, *105*, 11326.

(101) Truhlar, D. G.; Gao, J.; García-Viloca, M.; Alhambra, C.; Corchado, J.; Luz Sánchez, M.; Poulsen, T. D. *Int. J. Quant. Chem.* **2004**, *100*, 1136.

(102) Nam, K.; Prat-Resina, X.; García-Viloca, M.; Devi-Kesavan, L. S.; Gao, J. *J. Am. Chem. Soc.* **2004**, *126*, 1369.

(103) Pu, J.; Gao, J.; Truhlar, D. G. *Chem. Rev.* **2006**, *106*, 3140.

(104) Pang, J.; Pu, J.; Gao, J.; Truhlar, D. G.; Allemann, R. K. *J. Am. Chem. Soc.* **2006**, *128*, 8015.

(105) Scrutton, N. S.; Basran, J.; Sutcliffe, M. J. *Eur. J. Biochem.* **1999**, *264*, 666.

(106) Antoniou, D.; Caratzoulas, S.; Kalyanaraman, C.; Mincer, J. S.; Schwartz, S. D. *Eur. J. Biochem.* **2002**, *269*, 3103.

(107) Nagel, Z. D.; Klinman, J. P. *Nat. Chem. Biol.* **2009**, *5*, 543.

(108) Lamzin, V. S.; Dauter, Z.; Popov, V. O.; Harutyunyan, E. H.; Wilson, K. S. *J. Mol. Biol.* **1994**, *236*, 759.

(109) Tishkov, V. I.; Popov, V. O. *Biochemistry (Moscow)* **2004**, *69*, 1252.

(110) Mesentsev, A. V.; Lamzin, V. S.; Tishkov, V. I.; Ustinnikova, T. B.; Popov, V. O. *Biochem. J.* **1997**, *321*, 475.

(111) Weckbecker, A.; Hummel, W. *Biotechnol. Lett.* **2004**, *26*, 1739.

(112) Ernst, M.; Kaup, B.; Müller, M.; Bringer-Meyer, S.; Sahm, H. *Appl. Microbiol. Biotechnol.* **2005**, *66*, 629.

(113) Rojkova, A. M.; Galkin, A. G.; Kulakova, L. B.; Serov, A. E.; Savitsky, P. A.; Fedorchuk, V. V.; Tishkov, V. I. *FEBS Lett.* **1999**, *445*, 183.

(114) Castillo, R.; Oliva, M.; Martí, S.; Moliner, V. *J. Phys. Chem. B* **2008**, *112*, 10012.

(115) Roca, M.; Oliva, M.; Castillo, R.; Moliner, V.; Tuñón, I. *Chem. Eur. J.* **2010**, *16*, 11399.

(116) Wu, Y. D.; Houk, K. N. *J. Org. Chem.* **1993**, *58*, 2043.

(117) Schiøtt, B.; Zheng, Y. J.; Bruice, T. C. *J. Am. Chem. Soc.* **1998**, *120*, 7192.

(118) Wu, Y. D.; Lai, D. K. W.; Houk, K. N. *J. Am. Chem. Soc.* **1995**, *117*, 4100.

(119) Xue, H.; Wu, X.; Huskey, W. P. *J. Am. Chem. Soc.* **1996**, *118*, 5804.

(120) Hermes, J. D.; Morrical, S. W.; O'Leary, M. H.; Cleland, W. W. *Biochemistry* **1984**, *23*, 5479.

(121) Blanchard, J. S.; Cleland, W. W. *Biochemistry* **1980**, *19*, 3543.

(122) Iida, M.; Kitamura-Kimura, K.; Maeda, H.; Mineki, S. *Biosci. Biotechnol. Biochem.* **1992**, *56*, 1966.

(123) Tishkov, V. I.; Galkin, A. G.; Egorov, A. M. *Doklady Akademii nauk SSSR* **1991**, *317*, 745.

(124) Bandaria, J. N.; Cheatum, C. M.; Kohen, A. *J. Am. Chem. Soc.* **2009**, *131*, 10151.

(125) Borgis, D. C.; Lee, S.; Hynes, J. T. *Chem. Phys. Lett.* **1989**, *162*, 19.

(126) Borgis, D. C.; Hynes, J. T. *J. Phys. Chem.* **1996**, *100*, 1118.

(127) Thompson, W. H. *J. Chem. Phys.* **2003**, *118*, 1059.

(128) Colbert, D. T.; Miller, W. H. *J. Chem. Phys.* **1992**, *96*, 1982.

Chapter 5
Conclusions

In this Thesis we have implemented two different solvent coordinates with the aim of deepening into the origin of enzyme catalytic power. The use of solvent coordinates has proven to be a useful strategy to dissect the role of structure, flexibility and dynamics of the protein environment in catalysis and has led us to the conclusions that we recapitulate here.

1. Taking the electrostatic potential as a solvent coordinate allowed us to understand the evolution that the environment experiences during the enzymatic process of three different systems (*Xanthobacter autotrophicus* haloalkane dehalogenase, rat liver catechol O-methyltransferase and *Escherichia coli* dihydrofolate reductase).

2. The analysis of the frequencies, force constants and effective masses associated with the solvent coordinate shows that most of

the environmental motions participating in the electrostatic reorganization are slower compared to those in the solute. Therefore, a large fraction of the environmental motions, happening from the Michaelis complex to the Transition State, precedes solute motions during the chemical step

3. The comparison between the FESs obtained in the aqueous solution and the ones obtained in the enzyme using the electrostatic coordinate has shown that, even though both TS ensembles are located at similar values of the electrostatic potential coordinate, the RS in the aqueous solution is found further away along that coordinate from the TS than in the enzyme. These results indicate that the catalytic efficiency of the enzyme is related to the preorganization of the protein environment, much more adequate to accommodate the TS structure.

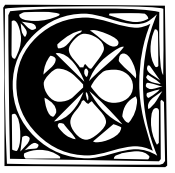
4. In order to measure the effect of an explicit consideration of the solvent coordinates in the definition of the TS ensemble, transmission coefficients and commitor histograms were calculated in two ways: considering only a solute coordinate and considering both the solute and solvent coordinates. Our results showed a significant improvement in the quality of the TS ensemble when the solvent coordinate is explicitly considered for reactions in an aqueous solution, which agrees with an increased participation of environmental motions in the uncatalyzed reaction.

5. We have implemented an energetic coordinate for QM/MM simulations, which permits the quantum treatment of primary and secondary hydrogen atoms. This coordinate was applied to study the hydride transfer reaction on the NAD-dependent formate dehydrogenase from the *methylotrophic bacterium Pseudomonas sp. 101*.

6. The TS defined according to this coordinate displays a wide range of donor-acceptor distances at which the hydride transfer can take place. Consequently, we developed a treatment where the hydride transfer was described in the adiabatic and non-adiabatic regimes.

7. The movement of the transferred hydrogen shows a strong coupling with the bending of the hydrogen in the acceptor atom, which is reflected in an increased secondary KIE.

8. Because of the small coupling at large distances the non-adiabatic regime contributes moderately to the observed rate constant in FDH. This is also supported by the calculated primary KIEs in both regimes. However, the calculated secondary KIEs indicate that the tunneling hydride transfer regime is not completely negligible.



hapter 6

Articles

ARTICLES
PUBLISHED ONLINE: 26 MAY 2013 | DOI: 10.1038/NCHEM.1660
nature chemistry

Studying the role of protein dynamics in an S_N2 enzyme reaction using free-energy surfaces and solvent coordinates

Rafael García-Meseguer[1], Sergio Martí[2], J. Javier Ruiz-Pernía[2]*, Vicent Moliner[2] and Iñaki Tuñón[1]*

Conformational changes are known to be able to drive an enzyme through its catalytic cycle, allowing, for example, substrate binding or product release. However, the influence of protein motions on the chemical step is a controversial issue. One proposal is that the simple equilibrium fluctuations incorporated into transition-state theory are insufficient to account for the catalytic effect of enzymes and that protein motions should be treated dynamically. Here, we propose the use of free-energy surfaces, obtained as a function of both a chemical coordinate and an environmental coordinate, as an efficient way to elucidate the role of protein structure and motions during the reaction. We show that the structure of the protein provides an adequate environment for the progress of the reaction, although a certain degree of flexibility is needed to attain the full catalytic effect. However, these motions do not introduce significant dynamical corrections to the rate constant and can be described as equilibrium fluctuations.

One of the most intriguing characteristics of enzymes is their flexibility. It has been stressed that, to function, enzymes must be stable enough to retain their three-dimensional (3D) structure, but flexible enough to permit the evolution of the protein among the different conformations relevant at each step of the full catalytic process[1,2]. For example, in many cases conformational changes are known to be a requisite for substrate binding and product release. In fact, the multidimensional free-energy surface that corresponds to an enzymatic process is very rugged, and contains multiple minima that appear along the multiple conformational coordinates available for the protein[3]. Transitions between different conformational substrates of the macromolecule and their population distribution are governed by motions that happen in a broad range of timescales, from milliseconds to femtoseconds[2,4].

However, the impact of protein flexibility on the rate constant of the chemical step remains the subject of a long-standing debate in scientific literature[5–15]. Even if the active site is designed to catalyse the reaction and then to accommodate the charge distribution of the transition state (TS)[16], some protein motions are still needed to evolve from the reactant state (RS) to the TS[1,5,15,17]. In other words, a certain degree of protein reorganization is needed to attain the TS[15]. This reorganization would play a similar role to that of solvent polarization in the Marcus theory for electron transfer in solution[18], although much slower conformational components could be involved in enzymatic catalysis[12,19]. In general, the coordinate that describes the transformation of the system from reactants to products is a collective coordinate that involves the degrees of freedom not only of the solute/substrate, but also of the environment (the solvent and/or the enzyme).

Theoretical descriptions of chemical reactions in enzymatic environments are usually based on the selection of a distinguishing reaction coordinate, typically defined in terms of the substrate or solute coordinates (that is, some valence coordinates related to the bonds to be broken and/or formed). Within this description, some protein motions can be coupled to the progress of the system along the reaction coordinate. For example, correlated motions within the protein promote tunnelling in hydride-transfer reactions[20,21] and compressive local motions within the active site facilitate the approach between the donor and the acceptor atoms in transfer reactions[11,22]. The participation of protein motions in the progress of the reaction is thus a well-established fact invoked using different terminology: protein reorganization[15], coupled motions[9] or promoting vibrations[8,22]. A more controversial question is the way in which these motions must be described and whether the current theoretical frameworks are adequate or not to explain the rate of enzymatic reactions[14,22].

As a fundamental approach to describe the reaction rate of chemical reactions, transition state theory (TST) provides the tools for analysing the rate of enzymatic reactions[23]. The basic assumption in TST is that the selected reaction coordinate (ζ) is separable from the rest of the coordinates of the system such that the averaged dynamic behaviour of the trajectories that evolve from the reactants to products can be represented as the equilibrium flux across the dividing surface. In this case the rate constant can be related simply to the free-energy difference between the TS and the reactants ($\Delta G^{\ddagger}$). For a unimolecular process

$$k_r = \frac{k_B T}{h} e^{-\frac{\Delta G^{\ddagger}(\zeta)}{RT}} \quad (1)$$

where k_B, h and R are the Boltzmann, Planck and gas constants, and T is the absolute temperature. This approach means that the remaining degrees of freedom can be considered at equilibrium at any value of the reaction coordinate, described as a Boltzmann distribution. If the reaction coordinate is defined exclusively in terms of the degrees of freedom of the substrate (or 'solute coordinate'), then the protein degrees of freedom (or 'solvent coordinate') are assumed to be at equilibrium at all the stages of the reaction path. Obviously, protein dynamics spans a hierarchy of timescales[2,4,10] and only those motions much faster than the reaction coordinate can be

[1]Departament de Química Física, Universitat de València, 46100 Burjassot, Spain, [2]Departament de Química Física i Analítica; Universitat Jaume I, 12071 Castellón, Spain. *e-mail: ignacio.tunon@uv.es; pernia@qfa.uji.es

NATURE CHEMISTRY DOI: 10.1038/NCHEM.1660

considered to be at equilibrium. Thus, this approach breaks down for motions coupled to the solute coordinate and that take place in similar or slower timescales. In such a case an explicit treatment of vibrational dynamics in the enzyme may be required. This is suggested for some cases covering both fast vibrational motions[22] and slow conformational changes[19]. However, other analyses stress that the role of protein motions can be incorporated satisfactorily in the description of the chemical step as equilibrium fluctuations and thus explicit dynamic treatments of protein motions are not really necessary for modelling enzymatic catalysis[5,9,13,14,24].

TST offers a convenient framework to incorporate non-statistical effects of protein motions. A transmission coefficient (κ) can be calculated from time-dependent simulations and incorporated into the TST expression of the rate constant[25]:

$$k_r = \kappa(\zeta)\frac{k_B T}{h} e^{-\frac{\Delta G^{\ddagger}(\zeta)}{RT}} \quad (2)$$

As the transmission coefficient takes values lower than unity the equilibrium approach (equation (1)) provides an upper limit to the rate constant. There is no unambiguous way to separate the effects of the enzyme on the activation free energy (or on the equilibrium) from those on the transmission coefficient (non-equilibrium effects) because both depend on the choice of the reaction coordinate[23]. If this is defined exclusively on the basis of the coordinates of the substrate, deviations from the equilibrium distribution of protein motions should be reflected in the transmission coefficient. Assuming that these effects are important in the vicinity of the dynamic bottleneck, the transmission coefficient can be evaluated from the frictional force exerted on the reaction coordinate at the TS[26] or by means of rare-event simulations that count the number of recrossings across the dividing surface[27]. To date, simulations performed on enzymatic reactions show that the transmission coefficient deviates only modestly from unity for reasonable choices of the reaction coordinate. Typical values usually range between 0.5 and 0.9 (refs 23,24,28), which reflects a modest participation of protein motions in the passage of the system over the barrier top.

A convenient way to analyse quantitatively the role of protein motions during the chemical transformation is to project the multidimensional free-energy surface (FES) of the enzymatic reaction in a 2D model obtained as a function of a solute coordinate and a solvent coordinate[29,30]. Such a FES allows one to estimate, at a quantitative level, the timing and coupling between the solute and solvent motions along the reaction path. Furthermore, the transmission coefficient can be derived from the differences between the dividing surface defined on the FES obtained under the equilibrium solvation approximation and that obtained on the 2D FES (which reflects the coupling between the solute and solvent coordinates).

A prototypical example in which solvent effects can play an important role is the S_N2 reaction. Studies in aqueous solution have already established that the reaction proceeds with a large rearrangement of water molecules around the nucleophile and the leaving group[30,31]. Thus, the change in the electrostatic potential created by the environment on the nucleophile and the leaving group is a good coordinate on which to follow the evolution of the environment along the reaction process[32], and the distances associated with the bonds to be broken and formed provide an adequate solute coordinate. In this work, we studied the S_N2 nucleophilic reaction between dichloroethane and Asp124 in DhlA, a haloalkane dehalogenase from *Xanthobacter autotrophicus* (Fig. 1)[33,34], an enzymatic reaction analysed theoretically by several groups[35–38]. We also modelled the counterpart S_N2 reaction in aqueous solution, in which a molecule of acetate is employed as the nucleophile. The comparison between the enzymatic and in-solution process offers an excellent opportunity to analyse the role of structure, flexibility and dynamics of the environment on a fundamental class of chemical reactions.

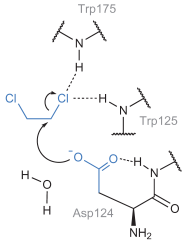


Figure 1 | Schematic representation of the S_N2 reaction catalysed by DhlA. The QM subsystem is depicted by the atoms labelled in blue.

Results

The PM3/MM FESs that correspond to nucleophilic attack in the aqueous solution and in the active site of DhlA traced as a function of the solute (ξ) and solvent (s) coordinates are presented in Fig. 2a,b, respectively. The solute coordinate ($\xi = d(\mathrm{CCl}) - d(\mathrm{CO})$) evolves from negative to positive values as the reaction proceeds. The solvent coordinate is obtained from the electrostatic potential created by the environment on the leaving group and the nucleophile ($s = V_{Cl} - V_O$; see Methods) and evolves from negative values at the RS (where the nucleophile atom bearing the negative charge is stabilized by electrostatic interactions with the environment) to positive values at the product state (where the electrostatic potential takes larger positive values on the leaving group that now is negatively charged). This figure also displays, as continuous lines on the FESs, the minimum free-energy paths (MFEPs) obtained from the free-energy gradient. The free-energy paths obtained using the solute coordinate and assuming that the solvent coordinate equilibrates at each value of the former (equilibrium free-energy paths (EFEPs)) are also shown in Fig. 2 (dashed lines).

Analysis of FESs. The free-energy differences between the saddle points and the reactant minima located on the FESs are 27.4 and 37.1 kcal mol^{-1} for the catalysed and uncatalysed processes, respectively. The activation free energy deduced from the experimental rate constant of the enzymatic reaction at 298 K is 15.3 kcal mol^{-1} (ref. 34) and the barrier estimated for the process in aqueous solution is 26 kcal mol^{-1} (ref. 36) (a value of 29.9 kcal mol^{-1} was reported for the reaction in solution at 373K)[35,39]. The overestimation observed in our theoretical values results from the use of the PM3 Hamiltonian[35,37]. The M06-2X-corrected free-energy barriers (see Methods) for the enzymatic and in-solution processes are 16.5 and 27.4 kcal mol^{-1}, respectively, in better quantitative agreement with the experimental values. In any case, the PM3/MM calculations provide a correct estimation of the catalytic effect, defined as the difference between the in-solution and the enzymatic free-energy barriers. The PM3/MM difference is 9.7 kcal mol^{-1}, in good agreement with the difference derived from the M06-2X values, 10.9 kcal mol^{-1}, and from the experimental values, 10.7 kcal mol^{-1}.

The FESs show noticeable differences between the reaction in the solution and in the enzyme. For the RS the protein structure provides a much more adequate environment for the progress of the reaction than does the solution. In solution, the RS occurs at a value for the solvent coordinate of about -100 kcal mol^{-1} |e|$^{-1}$, and the enzymatic RS occurs at about -25 kcal mol^{-1} |e|$^{-1}$. This latter value of the solvent coordinate is much closer to the value needed to reach the TS

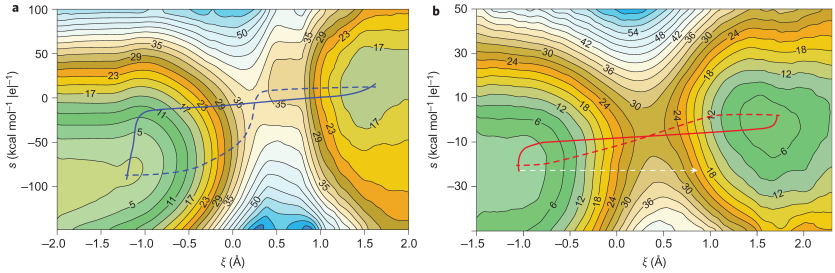


Figure 2 | FESs corresponding to the S_N2 reaction between an acetate anion and dichloroethane. a,**b**, FESs for the reaction in aqueous solution (**a**) and in the active site of DhlA (**b**). For **a** and **b** the solute distinguishing coordinate is the antisymmetric combination of the bond-breaking and -forming distances ($\xi = d(\mathrm{ClC}) - d(\mathrm{OC})$); the solvent coordinate is the antisymmetric combination of the electrostatic potential created by the environment on the leaving group and the nucleophilic oxygen ($s = V_{Cl} - V_O$). The isoenergetic free-energy lines represent steps of 3 kcal mol^{-1}. The continuous red and blue lines represent the minimum free-energy path on the FESs and the dashed ones correspond to the path obtained assuming equilibrium solvation along the solute coordinate. The white arrow represents the reaction path from the enzymatic Michaelis complex in a completely rigid protein environment.

($s \approx -10$ kcal mol^{-1} |e|$^{-1}$ in both environments). For the Michaelis complex the protein is already organized, from the electrostatic point of view, to favour the reaction, but in aqueous solution the environment largely needs to be reorganized to facilitate the reaction[30,31,38].

Nevertheless, the protein structure does not behave as a rigid scaffold in which the reaction takes place. The reaction would be significantly more difficult in a frozen-protein environment in which the solvent coordinate remains unchanged from RS to TS. A straight line from the enzymatic RS at a constant value of the solvent coordinate represents the rigid-environment path. This path (white arrow in Fig. 2b) shows a PM3/MM free-energy barrier of 31.5 kcal mol^{-1}, ~4 kcal mol^{-1} higher than the value observed for the flexible protein. A rate constant that corresponds to a hypothetical rigid protein would be about 10^3 times smaller, at room temperature, than that in the real enzyme. This means that protein flexibility plays a role in catalysis and that the environment needs to be rearranged when going from the Michaelis complex to the TS to reach a maximum reduction in the activation free energy.

Flexibility can be quantified by means of the force constant associated with the solvent coordinate. The force constants obtained from a parabolic fit of the free-energy change along the solvent coordinate in solution and in DhlA are given in Table 1. The force constant obtained in the enzyme is about 4.2 times larger than that obtained in the solution. The protein structure is stiffer than the structure of water, which is related to the existence of a network of covalent bonds in the former. However, as stated above, the change needed in the solvent coordinate to reach the TS from the RS is much smaller in the enzyme than in the solution. The final result is that the work to be done on the solvent coordinate to reach the TS is significantly smaller in the enzyme than in the solution. According to the MFEP traced on the FESs, the free-energy difference between the TS and the RS can be written as approximately

$$\Delta G_i^{\ddagger} \approx \Delta G_{s,i}(s_i^{\mathrm{RS}} \rightarrow s_i^{\ddagger}; \xi_i^{\mathrm{RS}}) + \Delta G_{\xi,i}(\xi_{\mathrm{enz}}^{\mathrm{RS}} \rightarrow \xi_{\mathrm{enz}}^{\ddagger}; s_i^{\ddagger}) \qquad (3)$$

where i stands for the environment (enzyme or the water solution). The first term of the right-hand side of equation (3) represents the work to be done on the solvent coordinate and the second one the work to be done on the solute coordinate. The values that correspond to the solvent coordinate are about 11 and 3 kcal mol^{-1} in aqueous solution and in DhlA, respectively. Thus, the free-energy cost associated with the change along the solvent coordinate in the enzyme is substantially smaller than that in the solution, and this difference represents 80% of the catalytic effect. It is interesting that, in the enzyme, the probability of sampling configurations of the environment conductive to the reaction ($s^{\ddagger}$) is much larger than the probability of sampling adequate values of the solute coordinate ($\xi^{\ddagger}$). The work associated with the reorganization of the environment in the enzymatic process represents only 11% of the total free-energy barrier.

Another important aspect to be analysed is the timing between the solute and solvent motions. According to the MFEPs obtained in the two environments, the change in the solvent coordinate precedes the change in the solute coordinate. From a dynamic point of view this means that the solvent coordinate is slower than the solute coordinate. A significant difference in the timescales of these two motions could result in important dynamical effects because of the delay between them. To characterize the time evolution of the environment we computed the characteristic frequencies associated with the motion along the solvent coordinate in the two environments using force constants and effective masses deduced from the equipartition principle. These frequencies are provided in Table 1. As observed, both the force constant and the effective mass associated with the solvent coordinate are larger in the enzyme than in the solution. Both effects cancel out and, as a result, the frequency associated with the motion along the solvent coordinate in the enzyme (410 cm^{-1}) is very similar to the value obtained in the solution (480 cm^{-1}). These values essentially correspond to the reorientation of hydrogen-bond donors around the nucleophile and the leaving group, motions that occur in picoseconds or faster[38]. So, from the dynamical point of view there are no significant differences in the participation of the environment during the reaction progress in aqueous solution or in DhlA. According to the MFEPs, the motions along the solvent coordinate, breaking of hydrogen bonds to the nucleophile and forming of new ones to the leaving group occur before or after the motion along the solute coordinate and the timescales associated

Table 1 | Force constants, effective masses and characteristic frequencies associated with the solvent coordinate (s) for the S_N2 reaction in solution and in DhlA.

	Aqueous solution	DhlA
K_s (kcal^{-1} mol |e|^2)	3.3×10^{-3}	1.4×10^{-2}
m_s (kcal^{-1} mol |e|^2 s^2)	0.40×10^{-30}	2.30×10^{-30}
ν_s (cm^{-1})	480	410

NATURE CHEMISTRY DOI: 10.1038/NCHEM.1660 ARTICLES

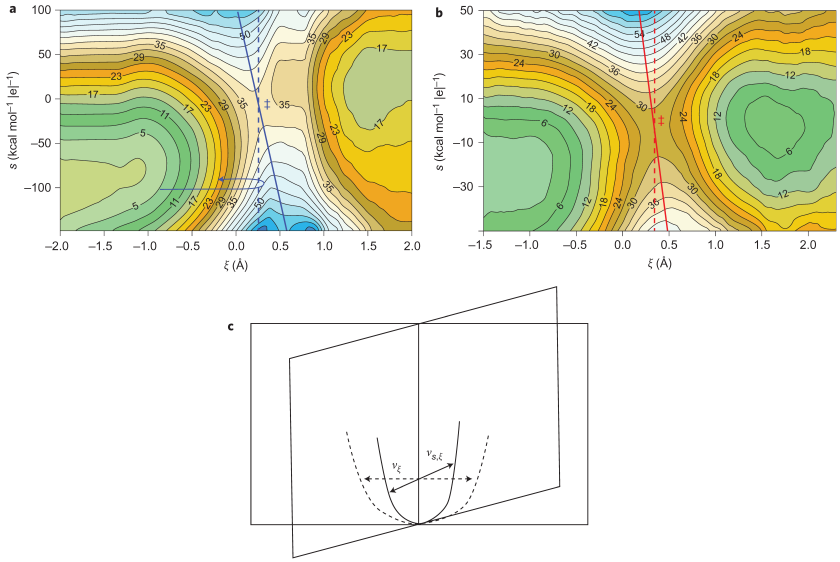


Figure 3 | The TS ensembles obtained from equilibrium and non-equilibrium pictures differ slightly, as reflected in the respective dividing surfaces. **a**, Dividing surfaces for the reaction in the solution. The arrow represents a hypothetical trajectory recrossing the equilibrium dividing surface. **b**, Dividing surfaces for the reaction in the active site of DhlA. The angle between the equilibrium and non-equilibrium plane is smaller than that in the solution. **c**, Schematic representation of the TS wells projected on the equilibrium and non-equilibrium dividing surfaces with their characteristic frequencies (ν_ξ and $\nu_{s,\xi}$). For **a**, **b** and **c** the continuous lines correspond to the dividing surface defined according to the FES traced along the solute and solvent coordinates. The dashed lines correspond to the dividing surface obtained when the solvent coordinate is assumed to be at equilibrium (the equilibrium plane is then defined just as $\xi = \xi^{\ddagger}$).

with these movements are very similar in the two media. Interestingly, we did not find any evidence that slow conformational motions of the protein affected the chemical step. Obviously, these motions can exist, but either they do not have consequences on the electrostatic coordinate (and thus on the energetics of the reaction) or they do not happen in the close neighbourhood of the Michaelis complex in the free-energy landscape.

Once analysed, we can compare the results obtained using both the solute and the solvent coordinates with those obtained assuming equilibrium solvation at any value of the solute coordinate. This is the usual approximation employed to analyse chemical reactions in condensed environments. The EFEPs, also presented in Fig. 2, go from the reactant to the product minima through the saddle point. Then, because the free energy is a state function, 1D profiles traced along the solute coordinate provide almost the same activation free energies as those of 2D surfaces obtained as a function of the solute and solvent coordinates (the origin of the small differences is discussed below). However, although the free-energy differences are correct, the equilibrium solvation approach is unable to describe properly the timing between the solute and solvent coordinates. Effectively, as observed in Fig. 2, in the equilibrium treatment one pulls along the solute coordinate and the solvent coordinate abruptly changes in the vicinity of the TS, but in the MFEPs, both in the solution and in the enzyme, solvent motions precede the changes along the solute coordinate. In any case, this limitation does not affect the estimation of the reaction-rate constant, because this is determined mostly by the free-energy difference between the TS and the RS.

Evaluation of the transmission coefficient. Although the MFEPs and EFEPs coincide at the RS and TS, there is a small difference in the activation free energies estimated from the 2D (or non-equilibrium) and the 1D (or equilibrium) treatments. The origin of this difference is in the definition of the TS ensemble obtained in each treatment. As shown in Fig. 3, the dividing surface in the 2D description contains the saddle point and goes through the ridges that separate the reactant and product valleys. In the 1D treatment, only the solute coordinate is employed and thus the dividing surface is defined simply as $\xi = \xi^{\ddagger}$ (represented as dashed lines in Fig. 3). The difference corresponds to a rotation of the dividing surface in variational TST[40]. The TS further along the non-equilibrium dividing surface is narrower than the TS further along the equilibrium dividing surface (see Fig. 3c) and so the frequencies associated with the motion of the TS along the former dividing surface ($\nu_{s,\xi}$) are larger than those for the latter (ν_ξ). This can be translated into an entropic difference in the TS ensembles and consequently in the activation free energies[29]:

$$\Delta\Delta G^{\ddagger} = \Delta G^{\ddagger}(s,\xi) - \Delta G^{\ddagger}(\xi) = RT \ln \frac{\nu_{s,\xi}}{\nu_\xi} \qquad (4)$$

From our FESs we estimated that the ratios between frequencies are about 1.3 and 1.1; using these values we estimated that $\Delta\Delta G^{\ddagger}$

at 298 K is about 0.2 kcal mol^{-1} and 0.1 kcal mol^{-1} in the solution and in the enzyme, respectively. These free-energy differences are quite small, below the statistical uncertainty of typical free-energy simulations, and thus non-equilibrium effects make a very small contribution to the activation free energies.

Obviously, a smaller activation free energy is translated into a larger rate constant. This difference should then be compensated by the consideration of a transmission coefficient smaller than unity (see equation (2)). The origin of this value can be understood in terms of the dividing surfaces obtained in the non-equilibrium and the equilibrium descriptions. The dividing surface obtained from the non-equilibrium treatment is defined such that any trajectory arriving at that surface from the reactant side will continue to the product region because the free energy continuously decreases in that direction, and so the transmission coefficient is equal to unity for that surface. However, using the equilibrium dividing surface, some trajectories that go from reactants to products find a free-energy barrier after crossing this surface and they could return to the reactant side (see Fig. 3a). This means that, in this description, the transmission coefficient would be lower than unity. Obviously, using equation (2) the final estimation of the rate constant obtained from the non-equilibrium and the equilibrium approaches would be the same if the transmission coefficient is obtained as[29]

$$\kappa = e^{-\frac{\Delta\Delta G^{\ddagger}}{RT}} \tag{5}$$

Using the free-energy differences given above, the transmission coefficients for the reaction in the solution and in the enzyme obtained using the solute coordinate as the distinguishing reaction coordinate are 0.8 and 0.9, respectively. Obviously, this procedure leads to a very crude estimate of the transmission coefficient because of the statistical errors associated with the free energies and of the non-explicit treatment of all the degrees of freedom. A more accurate estimation shows that the transmission coefficients for this reaction are about 0.6 and 0.8 in aqueous solution and in the enzyme, respectively[38]. The transmission coefficients, evaluated at the TS, depend on the participation of the solvent coordinate in the barrier-crossing event, which is quite small for this reaction. However, it must be stressed again that this does not mean that the protein remains unchanged during the chemical step. Protein motions occur first and facilitate the motion along the solute coordinate, but at the TS they can be considered to be in equilibrium. In other words, although the reactions involve protein motions coupled to the chemical transformation, the probability that these motions take the system to the TS is determined mainly by the activation free energy[1,5,15].

Discussion

The use of an explicit solvent coordinate can be a useful strategy to dissect the role of structure, flexibility and dynamics in catalysis. Comparison of the FESs obtained in the solution and in the enzyme for a prototypic S_N2 reaction shows that the origin of the catalytic efficiency of the enzyme is due to a protein structure that provides an adequate environment for the progress of the reaction. However, the protein is not completely rigid and some motions contribute to reduce the free-energy barrier. These motions, which occur on a picosecond timescale or faster, dynamically precede the change in the solute coordinate. We also show that these protein motions can be treated as equilibrium fluctuations with the rate constant determined mainly by the free-energy difference between the TS and the RS. The activation free energies estimated under the assumption that the solvent coordinate is in equilibrium with the solute coordinate are only slightly underestimated with respect to those obtained in the non-equilibrium description. These differences can be accounted for by the inclusion of a transmission coefficient in the TST rate constant.

We believe that our analysis of this reaction can be extended easily to other enzymatic systems and so provide an adequate framework for a quantitative discussion on the role of protein motions in catalysis.

Methods

We used a quantum mechanics/molecular mechanics (QM/MM) computational scheme in which dichloroethane and the side chain of residue Asp 124 are described using the PM3 semi-empirical Hamiltonian[41]. The rest of the enzyme and/or water molecules form the MM subsystem described by means of the all-atoms optimized potential for liquid simulation (OPLS) for the enzyme[42] and a flexible TIP3P potential for water molecules[43]. The Lennard–Jones parameters for the QM/MM interactions are also taken from the OPLS potential, except those for the chlorine atoms, which are taken from Gao and Xia[44].

For the system in aqueous solution we placed dichloroethane and acetate in a pre-equilibrated box of water molecules of side 55.8 Å, deleting all those water molecules with oxygen atoms found at <2.8 Å from any non-hydrogen atom of the solute fragments. For the enzyme-substrate system the X-ray crystal structure coordinates were taken from the Protein Data Bank (code 2DHC)[45]. The protonation state of titratable residues was determined using the PropKa program[46]. The whole system was placed in a pre-equilibrated cubic box of water molecules of side 79.5 Å. To neutralize the charge of the protein, 16 sodium ions were added so that both in the solution and in the enzymatic system the total charge was −1. The initial coordinates for the TSs in both environments are taken from our previous work[37,38]. The systems were equilibrated further by means of 200 ps of molecular dynamics simulation in the NVT ensemble at the reference temperature of 298 K using the Langevin integrator with a time step of 1 fs and periodic boundary conditions. A cutoff radii switched between 12.5 and 15 Å was applied for all types of interaction.

In this work we obtained FESs using two different coordinates, a solute coordinate (ξ) and a solvent coordinate (s). The FES can be expressed as

$$W(\xi, s) = C' - kT \ln \int \rho(x^N)\delta(\xi(x^N - \xi_0) - r)\delta(s(x^N) - s_0)dx^N \tag{6}$$

where $\rho(x^N)$ is the probability density of finding the system at the configuration x^N. In this case, the solute coordinate (ξ) is the antisymmetric combination of the distances of the outgoing chloride and the incoming oxygen to the carbon atom ($\xi = d(\mathrm{ClC}) - d(\mathrm{OC})$). The solvent coordinate selected was the antisymmetric combination of the electrostatic potential created by the environment on the outgoing chlorine atom and the incoming oxygen atom, the electrostatic potential created by the MM environment on the chlorine atom:

$$s = V_{\mathrm{Cl}}(x^N) - V_{\mathrm{O}}(x^N) = \sum_{j=1}^{M} \frac{q_j}{\left|x_j - x_{\mathrm{Cl}}\right|} - \sum_{j=1}^{M} \frac{q_j}{\left|x_j - x_{\mathrm{O}}\right|} \tag{7}$$

where the sums run on the M sites of the environment with point charges q_j. This is a collective coordinate that involves all the MM atoms with an electrostatic influence on the donor or acceptor atoms.

The FESs that correspond to the reactions in aqueous solution and the enzyme were obtained using umbrella sampling[47], applying parabolic constraints to the solute and the solvent coordinates:

$$V_r = \frac{1}{2}K_\xi(\xi - \xi_0)^2; \quad V_s = \frac{1}{2}K_s(s - s_0)^2 \tag{8}$$

Molecular dynamics preferentially explores the most probable configurations of the system around the reference values ξ_0 and s_0. The joint probability distributions of the two coordinates were obtained by means of the weighted histogram analysis method (WHAM)[48]. To save computational cost, simulations were performed with any atom beyond 25 Å of dichloroethane kept frozen. A total of 5,454 simulation windows consisting of 5 ps of equilibration and 50 ps of production were employed to trace the FESs in the solution, and for the enzyme 3,100 windows were needed. The force constants used to keep the system at the reference values of the solute and solvent coordinates were 2,500 kJ mol^{-1} Å^{-2} and 0.01 kJ $^{-1}$ mol $|e|^2$, which provided a good control of the coordinates[32].

Importantly, the PM3/MM Hamiltonian results in systematically overestimated energy barriers, but the geometries obtained for the RS and TS are good enough for reasonable estimations of kinetic isotope effects[35]. We then corrected the systematic error in the activation free energies by means of single-point calculations at higher theoretical levels. With this purpose we optimized ten TS structures starting from different configurations selected from the corresponding simulation. After intrinsic reaction coordinate calculation, the energy barrier was obtained at the PM3/MM level and by means of single-point calculations at the M06-2X/6-311 + G(2df,2p)/MM level[49]. The correction energy term was evaluated as the averaged difference between the semi-empirical and M06-2X energy barriers.

Received 15 January 2013; accepted 18 April 2013; published online 26 May 2013

References

1. Hammes, G. G., Benkovic, S. J. & Hammes-Schiffer, S. Flexibility, diversity, and cooperativity: pillars of enzyme catalysis. *Biochemistry* **50,** 10422–10430 (2011).
2. Henzler-Wildman, K. A. *et al.* A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **450,** 913–916 (2007).
3. Benkovic, S. J., Hammes, G. G. & Hammes-Schiffer, S. Free-energy landscape of enzyme catalysis. *Biochemistry* **47,** 3317–3321 (2008).
4. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450,** 964–972 (2007).
5. Garcia-Viloca, M., Gao, J., Karplus, M. & Truhlar, D. G. How enzymes work: analysis by modern rate theory and computer simulations. *Science* **303,** 186–195 (2004).
6. Gao, J. *et al.* Mechanisms and free energies of enzymatic reactions. *Chem. Rev.* **106,** 3188–3209 (2006).
7. Olsson, M. H. M., Parson, W. W. & Warshel, A. Dynamical contributions to enzyme catalysis: critical tests of a popular hypothesis. *Chem. Rev.* **106,** 1737–1756 (2006).
8. Antoniou, D., Basner, J., Núñez, S. & Schwartz, S. D. Computational and theoretical methods to explore the relation between enzyme dynamics and catalysis. *Chem. Rev.* **106,** 3170–3187 (2006).
9. Nashine, V. C., Hammes-Schiffer, S. & Benkovic, S. J. Coupled motions in enzyme catalysis. *Curr. Opin. Chem. Biol.* **14,** 644–651 (2010).
10. Ramanathan, A. & Agarwal, P. K. Evolutionarily conserved linkage between enzyme fold, flexibility, and catalysis. *PLoS Biol* **9,** e1001193 (2011).
11. Zhang, J. & Klinman, J. P. Enzymatic methyl transfer: role of an active site residue in generating active site compaction that correlates with catalytic efficiency. *J. Am. Chem. Soc.* **133,** 17134–17137 (2011).
12. Bhabha, G. *et al.* A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science* **332,** 234–238 (2011).
13. Adamczyk, A. J., Cao, J., Kamerlin, S. C. L. & Warshel, A. Catalysis by dihydrofolate reductase and other enzymes arises from electrostatic preorganization, not conformational motions. *Proc. Natl Acad. Sci. USA* **108,** 14115–14120 (2011).
14. Glowacki, D. R., Harvey, J. N. & Mulholland, A. J. Taking Ockham's razor to enzyme dynamics and catalysis. *Nature Chem.* **4,** 169–176 (2012).
15. Kamerlin, S. C. L. & Warshel, A. At the dawn of the 21st century: is dynamics the missing link for understanding enzyme catalysis? *Proteins* **78,** 1339–1375 (2010).
16. Warshel, A. *et al.* Electrostatic basis for enzyme catalysis. *Chem. Rev.* **106,** 3210–3235 (2006).
17. Kurplus, M & McCammon, J A. Dynamics of proteins: elements and function. *Annu. Rev. Biochem.* **52,** 263–300 (1983).
18. Marcus, R. A. Chemical and electrochemical electron-transfer theory. *Annu. Rev. Phys. Chem.* **15,** 155–196 (1964).
19. Kosugi, T. & Hayashi, S. Crucial role of protein flexibility in formation of a stable reaction transition state in an α-amylase catalysis. *J. Am. Chem. Soc.* **134,** 7045–7055 (2012).
20. Pang, J., Pu, J., Gao, J., Truhlar, D. G. & Allemann, R. K. Hydride Transfer reaction catalyzed by hyperthermophilic dihydrofolate reductase is dominated by quantum mechanical tunneling and is promoted by both inter- and intramonomeric correlated motions. *J. Am. Chem. Soc.* **128,** 8015–8023 (2006).
21. Kanaan, N. *et al.* Temperature dependence of the kinetic isotope effects in thymidylate synthase. A theoretical study. *J. Am. Chem. Soc.* **133,** 6692–6702 (2011).
22. Hay, S. & Scrutton, N. S. Good vibrations in enzyme-catalysed reactions. *Nature Chem.* **4,** 161–168 (2012).
23. Pu, J., Gao, J. & Truhlar, D. G. Multidimensional tunneling, recrossing, and the transmission coefficient for enzymatic reactions. *Chem. Rev.* **106,** 3140–3169 (2006).
24. Boekelheide, N., Salomón-Ferrer, R. & Miller, T. F. Dynamics and dissipation in enzyme catalysis. *Proc. Natl Acad. Sci. USA* **108,** 16159–16163 (2011).
25. Garrett, B. C. & Truhlar, D. G. in *Theory and Applications of Computational Chemistry* (eds Dykstra, C. E., Frenking, G., Kim, K. S. & Scuseria, G. E.) 67–87 (Elsevier, 2005).
26. Grote, R. F. & Hynes, J. T. The stable states picture of chemical reactions. II. Rate constants for condensed and gas phase reaction models. *J. Chem. Phys.* **73,** 2715–2732 (1980).
27. Bergsma, J. P., Gertner, B. J., Wilson, K. R. & Hynes, J. T. Molecular dynamics of a model S_N2 reaction in water. *J. Chem. Phys.* **86,** 1356–1376 (1987).
28. Ruiz-Pernía, J. J., Tuñón, I., Moliner, V., Hynes, J. T. & Roca, M. Dynamic effects on reaction rates in a Michael addition catalyzed by chalcone isomerase. Beyond the frozen environment approach. *J. Am. Chem. Soc.* **130,** 7477–7488 (2008).
29. Gertner, B. J., Bergsma, J. P., Wilson, K. R., Lee, S. & Hynes, J. T. Nonadiabatic solvation model for S_N2 reactions in polar solvents. *J. Chem. Phys.* **86,** 1377–1386 (1987).
30. Hwang, J. K., King, G., Creighton, S. & Warshel, A. Simulation of free energy relationships and dynamics of S_N2 reactions in aqueous solution. *J. Am. Chem. Soc.* **110,** 5297–5311 (1988).
31. Gertner, B. J., Wilson, K. R. & Hynes, J. T. Nonequilibrium solvation effects on reaction rates for model S_N2 reactions in water. *J. Chem. Phys.* **90,** 3537–3558 (1989).
32. Ruiz-Pernía, J. J., Martí, S., Moliner, V. & Tuñón, I. A novel strategy to study electrostatic effects in chemical reactions: differences between the role of solvent and the active site of chalcone isomerase in a Michael addition. *J. Chem. Theory Comput.* **8,** 1532–1535 (2012).
33. Janssen, D. B., Scheper, A., Dijkhuizen, L. & Witholt, B. Degradation of halogenated aliphatic-compounds by *Xanthobacter autotrophicus* GJ10. *Appl. Environ. Microbiol.* **49,** 673–677 (1985).
34. Schanstra, J. P., Kingma, J. & Janssen, D. B. Specificity and kinetics of haloalkane dehalogenase. *J. Biol. Chem.* **271,** 14747–14753 (1996).
35. Devi-Kesavan, L. S. & Gao, J. Combined QM/MM study of the mechanism and kinetic isotope effect of the nucleophilic substitution reaction in haloalkane dehalogenase. *J. Am. Chem. Soc.* **125,** 1532–1540 (2003).
36. Shurki, A., Štrajbl, M., Villà, J. & Warshel, A. How much do enzymes really gain by restraining their reacting fragments? *J. Am. Chem. Soc.* **124,** 4097–4107 (2002).
37. Soriano, A. *et al.* Electrostatic effects in enzyme catalysis: a quantum mechanics/molecular mechanics study of the nucleophilic substitution reaction in haloalkane dehalogenase. *Theor. Chem. Accounts* **112,** 327–334 (2004).
38. Soriano, A., Silla, E., Tuñón, I. & Ruiz-Lopez, M. F. Dynamic and electrostatic effects in enzymatic processes. An analysis of the nucleophilic substitution reaction in haloalkane dehalogenase. *J. Am. Chem. Soc.* **127,** 1946–1957 (2005).
39. Okamoto, K., Kita, T., Araki, K. & Shingu, H. Kinetic studies of bimolecular nucleophilic substitution. 4. Rates of S_N2 and E2 reactions of beta-substituted ethyl chlorides with sodium acetate in aqueous solutions. *B. Chem. Soc. Jpn* **40,** 1913 (1967).
40. Truhlar, D. G. & Garrett, B. C. Variational transition-state theory. *Acc. Chem. Res.* **13,** 440–448 (1980).
41. Stewart, J. J. P. Optimization of parameters for semiempirical methods I. Method. *J. Comput. Chem.* **10,** 209–220 (1989).
42. Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **118,** 11225–11236 (1996).
43. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79,** 926–935 (1983).
44. Gao, J. & Xia, X. A two-dimensional energy surface for a type II S_N2 reaction in aqueous solution. *J. Am. Chem. Soc.* **115,** 9667–9675 (1993).
45. Verschueren, K. H., Seljée, F., Rozeboom, H. J., Kalk, K. H. & Dijkstra, B. W. Crystallographic analysis of the catalytic mechanism of haloalkane dehalogenase. *Nature* **363,** 693–698 (1993).
46. Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: consistent treatment of internal and surface residues in empirical pK_a predictions. *J. Chem. Theory Comput.* **7,** 525–537 (2011).
47. Torrie, G. M. & Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. *J. Comput. Phys.* **23,** 187–199 (1977).
48. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13,** 1011–1021 (1992).
49. Zhao, Y. & Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Accounts* **120,** 215–241 (2008).

Acknowledgements

The authors acknowledge financial support from the Ministerio de Economía y Competitividad (MEC) through project CTQ2012-36253-C03. J.J.R-P. thanks a Juan de la Cierva contract and R.G-M. a FPU fellowship of the Ministerio de Economía y Competitividad. I.T. acknowledges helpful discussions held with D. Laage and J. T. Hynes during his sabbatical stay at the École Normale Supérieure, France. The authors acknowledge computational facilities of the Servei d'Informàtica de la Universitat de València on the 'Tirant' supercomputer.

Author contributions

I.T., V.M. and J.J.R-P. designed the computational experiments. S.M. wrote the code and R.G-M. performed the calculations. I.T., V.M. and J.J.R-P. co-wrote the first version of the paper. All the authors commented and discussed the results and the final version of the manuscript.

Additional information

Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.J.R-P. and I.T.

Competing financial interests

The authors declare no competing financial interests.

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

Increased Dynamic Effects in a Catalytically Compromised Variant of *Escherichia coli* Dihydrofolate Reductase

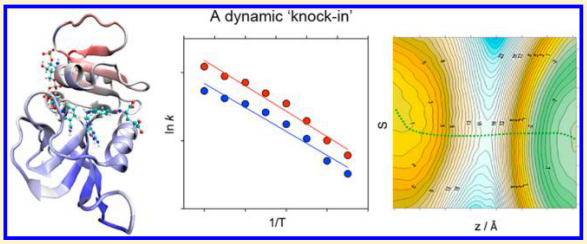
J. Javier Ruiz-Pernia,[†,∥] Louis Y. P. Luk,[‡,∥] Rafael García-Meseguer,[§] Sergio Martí,[†] E. Joel Loveridge,[‡] Iñaki Tuñón,*[,§] Vicent Moliner,*[,†] and Rudolf K. Allemann*[,‡]

[†]Departament de Química Física i Analítica, Universitat Jaume I, 12071 Castello, Spain

[‡]School of Chemistry & Cardiff Catalysis Institute, Cardiff University, Park Place, Cardiff, CF10 3AT, U.K.

[§]Departament de Química Física, Universitat de València, 46100 Burjassot, Spain

Supporting Information



ABSTRACT: Isotopic substitution (^{15}N, ^{13}C, ^{2}H) of a catalytically compromised variant of *Escherichia coli* dihydrofolate reductase, EcDHFR-N23PP/S148A, has been used to investigate the effect of these mutations on catalysis. The reduction of the rate constant of the chemical step in the EcDHFR-N23PP/S148A catalyzed reaction is essentially a consequence of an increase of the quasi-classical free energy barrier and to a minor extent of an increased number of recrossing trajectories on the transition state dividing surface. Since the variant enzyme is less well set up to catalyze the reaction, a higher degree of active site reorganization is needed to reach the TS. Although millisecond active site motions are lost in the variant, there is greater flexibility on the femtosecond time scale. The "dynamic knockout" EcDHFR-N23PP/S148A is therefore a "dynamic knock-in" at the level of the chemical step, and the increased dynamic coupling to the chemical coordinate is in fact detrimental to catalysis. This finding is most likely applicable not just to hydrogen transfer in EcDHFR but also to other enzymatic systems.

INTRODUCTION

The involvement of protein motions in the chemical step of enzyme reactions that involve hydrogen (H^+, $H^•$, or H^-) tunnelling remains one of the most discussed topics in modern enzymology. Early computational evidence suggested that dynamic effects may not be required to account for enzymatic catalysis and that the participation of protein motions in the chemical step can be satisfactorily described as equilibrium fluctuations.[1] Nevertheless, theoretical frameworks that involve protein "promoting vibrations" or "promoting motions" on femtosecond to millisecond time scales, which are proposed to reduce the height and/or width of the potential energy barrier and thereby enhance enzymatic catalysis,[2−4] have been invoked to interpret experimental data.[5−11] Protein motions occur over a hierarchy of time scales and ranges,[12,13] from femtosecond local bond vibrations to millisecond large scale domain motions. It has been suggested that motions on one time scale may facilitate motions on another.[14,15] A number of theoretical analyses based on atomistic simulations have suggested that protein motions may not significantly contribute to reducing the free energy barrier of enzyme catalyzed reactions,[16−20] and several experimental studies have shown that models which require a contribution from protein motions for catalysis do not seem fully compatible with the available data.[21−28] A combination of experimental results, QM/MM simulations, and theoretical analyses has however recently revealed that the dynamics of the protein environment do have a small but measurable effect on the chemical reaction.[28]

Dihydrofolate reductase (DHFR) has often been used in studies of the relationship between protein motions and catalysis.[5−7,29−33] It catalyzes the NADPH-dependent reduction of 7,8-dihydrofolate (H_2F) to 5,6,7,8-tetrahydrofolate (H_4F) by hydride transfer from C4 of NADPH and protonation of N5 of H_2F (Figure 1). The catalytic cycle of DHFR from *E. coli* (EcDHFR) has been thoroughly characterized;[33−35] the physical steps of substrate binding and product release involve large scale millisecond time scale

Received: October 17, 2013
Published: November 19, 2013

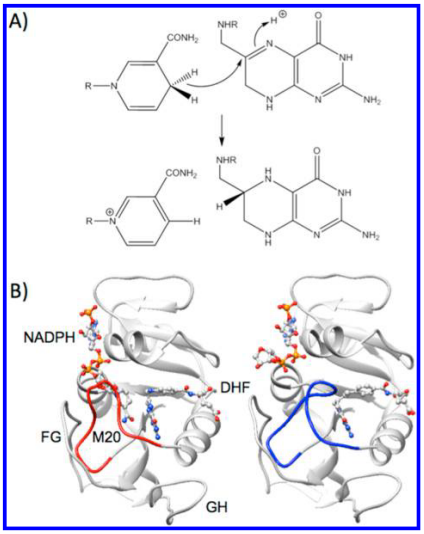


Figure 1. (A) Conversion of dihydrofolate to tetrahydrofolate through transfer of the pro-R hydride of NADPH and a solvent proton. (B) Cartoon representation of the closed and occluded conformations of EcDHFR. Catalytically important loops (M20, βFG, and GH), substrate (H_2F), and cofactor (NADPH) are labeled. The M20 loop is highlighted in red in the closed conformation and in blue in the occluded conformation. The cofactor and substrate are represented using ball-and-stick models.

conformational motions of the M20 loop (residues 9−24), which forms part of the active site.[33,34] The closed conformation, found in the holo-enzyme and the Michaelis complex,[34] is stabilized by hydrogen bonding from the M20 loop to the neighboring βFG loop (residues 116−132). Following hydride transfer, the M20 loop adopts the occluded conformation, which is stabilized by hydrogen bonds from the M20 loop to the neighboring βGH loop (residues 142−149). This prevents the nicotinamide ring of the cofactor NADP(H) from entering the active site[34] (Figure 1).

A recent report of a "dynamic knockout" of EcDHFR has reignited controversy over the role of protein motions in hydrogen transfer reactions.[36] EcDHFR-N23PP/S148A is unable to adopt the occluded conformation due to the absence of the crucial hydrogen bonds between the M20 and βGH loops, and millisecond to microsecond time scale motions observed in the M20 loop of wild type EcDHFR are lost in the variant.[36] This "dynamic knockout" displayed reduced hydride transfer rate constants, and it was proposed that the protein motions lost in EcDHFR-N23PP/S148A are involved in promoting hydride transfer in wild type EcDHFR.[36] However, theoretical studies based on the empirical valence bond approach suggested that the reduction in the hydride transfer rate constant in EcDHFR-N23PP/S148A was due to effects on the electrostatic preorganization and consequently the reorganization free energy within the active site.[37] Measurement of the kinetic isotope effects of the EcDHFR-N23PP/S148A catalyzed hydride transfer reaction provided further evidence that the loss of protein motions may not be the cause of the impaired catalysis.[27] A number of other studies have also suggested that protein motions do not promote the chemical step in wild type EcDHFR catalysis,[16,24−26,31,38] but the issue continues to generate debate in the literature.

The Born−Oppenheimer approximation assumes that, while isotopic substitutions do not affect the electronic potential energy surface, mass-dependent differences result in changes in atomic motions from femtosecond bond vibrations to millisecond time scale conformational changes.[39] Hence, enzymatic isotopic substitution to form "Born−Oppenheimer enzymes"[40,41] has been postulated to affect catalysis by changing protein motions that couple to the reaction coordinate.[40−43] Accordingly, we have recently shown by a combination of experimental and computational studies of light (natural isotopic abundance) and heavy (^{15}N, ^{13}C, ^{2}H isotopically substituted) EcDHFR that mass-dependent protein motions in EcDHFR affect the dynamic recrossing of hydride transfer but they do not promote tunnelling, and that the increased number of recrossing trajectories has a slight effect on the effective barrier of the chemical step.[28]

The present investigation analyzes the dynamic contributions to enzyme catalysis by comparing heavy EcDHFR-N23PP/S148A with its light counterpart. In particular, we report experimental and computational results that allow characterization of the dynamics of the coordinates associated with the hydride transfer from NADPH to H_2F. We provide strong evidence that the loss of protein motions observed previously by NMR[36] does not directly impact catalysis of the hydride transfer step; in fact, the chemical step is subject to *greater* dynamic contributions in the "dynamic knockout" than in the wild type enzyme. Together with the increased reaction barrier, the increased dynamic coupling to the chemical coordinate is the cause of the reduction of the hydride transfer rate observed in EcDHFR-N23PP/S148A.

■ RESULTS AND DISCUSSION

Creation of "Heavy" EcDHFR-N23PP/S148A. "Heavy" EcDHFR-N23PP/S148A was produced in M9 medium containing exclusively $^{15}NH_4Cl$, U−^{13}C,^{2}H-glucose, and $^{2}H_2O$. After purification in buffers made of $^{1}H_2O$, heavy EcDHFR-N23PP/S148A showed a 10.02% increase in molecular mass (Figures S1 and S2, Supporting Information), indicating that 91.2% of the ^{14}N, ^{12}C, and non-exchangeable ^{1}H atoms had been replaced by their heavier isotopes. The secondary structures of light and heavy EcDHFR-N23PP/S148A were indistinguishable as measured by circular dichroism spectroscopy (Figure S3, Supporting Information).

Experimental Results. The nature of the rate limiting step in DHFR catalysis is dependent on pH.[35] At pH 7, the release of $NADP^+$ from the binary EcDHFR-N23PP/S148A·$NADP^+$ complex is most likely rate limiting,[36] whereas the release of H_4F from the E·NADPH·H_4F mixed ternary complex is rate limiting in wild type EcDHFR.[35] Neither the Michaelis constants K_M nor the turnover numbers k_{cat} were sensitive to isotopic substitution of EcDHFR-N23PP/S148A (Tables S2 and S3, Supporting Information). In the EcDHFR·$NADP^+$ complex as well as in the apoenzyme, the loop regions are largely disordered[44] and hence $NADP^+$ release is unlikely to involve a significant conformational change in either the wild type or variant enzyme. In contrast, a significant enzyme KIE_{cat} ($k_{cat}^{LE}/k_{cat}^{HE}$, where LE and HE indicate light and heavy

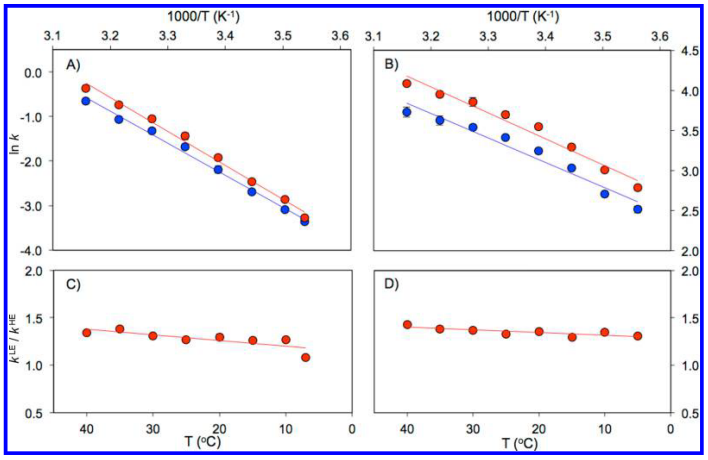


Figure 2. Temperature dependence of the experimental EcDHFR-N23PP/S148A hydride transfer rate constants. (A) pH 9.5 steady state kinetic data; (B) pH 7.0 pre-steady-state kinetic data. Data and Arrhenius fits are shown in red for the light enzyme and in blue for the heavy enzyme. (C and D) The enzyme KIE (ratio of light to heavy enzyme rate constants, k^{LE}/k^{HE}) at pH 9.5 and pH 7.0, respectively.

Table 1. Experimentally Determined Rate Constants and Enzyme KIEs for Hydride Transfer in Light and Heavy EcDHFR-N23PP/S148A at 25 °C and Activation Parameters from Fitting the Experimental Data to the Arrhenius Equation[a]

enzyme and pH	k (s^{-1})	enzyme KIE	E_a ($kJ \cdot mol^{-1}$)	ΔE_a ($kJ \cdot mol^{-1}$)	A_H (s^{-1})	A_H^{LE}/A_H^{HE}
light EcDHFR, pH 7	178.2 ± 4.7	1.10 ± 0.03	31.84 ± 0.69	5.78 ± 1.61	(6.42 ± 0.81) × 10^7	10.74 ± 0.13
heavy EcDHFR, pH 7	151.6 ± 4.2		26.05 ± 1.45		(5.98 ± 0.11) × 10^6	
light EcDHFR, pH 9.5	1.86 ± 0.18	1.13 ± 0.08	60.65 ± 0.72	3.15 ± 1.28	(7.49 ± 0.31) × 10^{10}	4.09 ± 0.24
heavy EcDHFR, pH 9.5	1.64 ± 0.16		57.50 ± 1.06		(1.83 ± 0.43) × 10^{10}	
light EcDHFR-N23PP/S148A, pH 7	40.32 ± 0.79	1.33 ± 0.02	27.14 ± 0.16	1.52 ± 0.31	(2.17 ± 0.14) × 10^6	2.50 ± 0.11
heavy EcDHFR-N23PP/S148A, pH 7	30.41 ± 0.80		25.62 ± 0.27		(8.63 ± 0.75) × 10^5	
light EcDHFR-N23PP/S148A, pH 9.5	0.24 ± 0.01	1.27 ± 0.03	63.98 ± 0.01	3.57 ± 0.79	(3.54 ± 0.51) × 10^{10}	5.44 ± 0.19
heavy EcDHFR-N23PP/S148A, pH 9.5	0.19 ± 0.01		60.41 ± 0.79		(6.51 ± 0.83) × 10^9	

[a]Data for light and heavy wild type EcDHFR are from ref 28.

enzyme, respectively) on the steady-state rate constants for wild type EcDHFR of 1.16 ± 0.01 at 35 °C was observed,[28] consistent with the large conformational change observed on H_4F release.[33,34]

For pH values above 9, hydride transfer from NADPH to H_2F determines k_{cat}.[27,35,36] The steady-state rate constants measured at pH 9.5 for light and heavy EcDHFR-N23PP/S148A showed a temperature dependent enzyme KIE_{cat}, such that k_{cat}^{LE} is 14% larger than k_{cat}^{HE} at 10 °C and 34% larger at 40 °C (Figure 2, Tables S1 and S2, Supporting Information). The enzyme KIE was higher in EcDHFR-N23PP/S148A than in wild type EcDHFR at all temperatures. In contrast, the Michaelis constants for NADPH and for H_2F are unaffected by isotopic substitution at both 20 and 35 °C (Table S3, Supporting Information), indicating that binding interactions are unchanged in the heavy enzyme and therefore that, at pH 9.5, the enzyme KIE reflects a difference in reactivity between the light and heavy enzymes after the formation of the respective Michaelis complexes.

Hydride transfer at pH 9.5 is not physiologically significant,[25] so the rate constants of the fast hydride transfer from reduced NADPH to mostly protonated H_2F were determined at pH 7.0 in pre-steady-state stopped-flow experiments, in which the fluorescence resonance energy transfer from the protein to reduced NADPH was measured. The pH dependence of the pre-steady-state hydride transfer rate constants indicated that the apparent pK_a of the EcDHFR-N23PP/S148A catalyzed reaction was not affected by isotopic substitution (Figure S4 and Table S4, Supporting Information). The apparent pK_a values for the reactions catalyzed by light and heavy EcDHFR-N23PP/S148A were 6.29 ± 0.10 and 6.23 ± 0.09 at 20 °C and 6.35 ± 0.16 and 6.15 ± 0.05 at 35 °C. The hydride transfer rate constants for light and heavy EcDHFR-N23PP/S148A (k_H^{LE} and k_H^{HE}) at pH 7.0 show a similar dependence on temperature to that observed in the steady state measurements at elevated pH (Table S1, Supporting Information; Figure 2). The enzyme KIE_H (k_H^{LE}/k_H^{HE}) increased from 1.31 ± 0.03 at 10 °C to 1.43 ± 0.04 at 40 °C (Figure 2). These are significantly higher than the corresponding values of 0.93 ± 0.02 at 10 °C to 1.18 ± 0.09 at 40 °C reported previously for the reaction catalyzed by wild type EcDHFR.[28] Activation energies were slightly higher in light EcDHFR-N23PP/S148A than in the heavy counterpart, with the reduction in free energy for the reaction resulting from lower Arrhenius prefactors and hence a less favorable activation

Table 2. Results from the QM/MM Simulations for Hydride Transfer in Light and Heavy EcDHFR-N23PP/S148A[a]

enzyme	γ	κ	ΔG_{act}^{QC} (kcal·mol^{-1})	ΔG_{eff} (kcal·mol^{-1})	k_{theor} (s^{-1})	$(k^{LE}/k^{HE})_{theor}$	k_H (s^{-1})	$(k^{LE}/k^{HE})_{exp}$
light EcDHFR-N23PP/S148A	0.53 ± 0.02	2.25 ± 0.45	16.43 ± 0.70	16.63 ± 0.84	8.0	1.26 ± 0.04	47.23 ± 1.28	1.37 ± 0.03
heavy EcDHFR-N23PP/S148A	0.42 ± 0.02			16.74 ± 0.84	6.3		34.44 ± 1.18	
light EcDHFR	0.57 ± 0.02	2.61 ± 0.49	14.59 ± 0.41	14.35 ± 0.54	219	1.16 ± 0.04	209.1 ± 5.0	1.10 ± 0.04
heavy EcDHFR	0.49 ± 0.02			14.46 ± 0.54	188		190.1 ± 8.5	

[a]Transmission coefficient components due to recrossing (γ) and tunnelling (κ), quasi-classical (QC) free energy of activation (ΔG_{act}^{QC}) (eq 1), effective phenomenological free energies of activation (ΔG_{eff}), and predicted (k_{theor}, at 300 K) and experimental (k_H, at 303 K) hydride transfer rate constants are included. 200 trajectories were obtained to give these data. Components of the quasi-classical activation free energy, the potential of mean force (PMF) difference between the TS and the reactants, and the classical and quantized vibration corrections are reported in the Supporting Information. Data for light and heavy wild type EcDHFR are from ref 28.

entropy (Table 1). The same situation had also been observed in the reaction catalyzed by wild type EcDHFR.[28] The reduced rate constants for hydride transfer in EcDHFR-N23PP/S148A relative to EcDHFR have also been shown to be the result of less favorable activation entropy in the variant, rather than increased activation enthalpies.[27]

Molecular Dynamics Simulations—Enzyme Isotope Effects. Our analysis of the rate constant of the chemical step is based on transition state theory (TST), modified to account for tunnelling contributions and other dynamic effects:[45−47]

$$k(T) = \Gamma(T,z)\frac{k_B T}{h}\exp\left(-\frac{\Delta G_{act}^{QC}(T,z)}{RT}\right) \quad (1)$$

where R is the ideal gas constant, T is the temperature, k_B is the Boltzmann constant, h is Planck's constant, $\Delta G_{act}^{QC}(T,z)$ is the quasi-classical activation free energy[48] (for details, see the Supporting Information) obtained as a function of a reaction coordinate z that corresponds to the transfer of the hydride from the donor atom of the cofactor to the substrate's acceptor atom ($z = d(C4_{NADPH}-H_t) - d(C6_{H2F}-H_t)$), and $\Gamma(T,z)$ is the temperature-dependent transmission coefficient, which contains the dynamic corrections to the classical rate constant and is therefore equal to unity in the limit of classical TST. Although different contributions are in principle coupled, $\Gamma(T,z)$ can be expressed as

$$\Gamma(T,z) = \gamma(T,z)\cdot\kappa(T) \quad (2)$$

where $\gamma(T,z)$ is the recrossing transmission coefficient that corrects the rate constant for trajectories that recross the dividing surface back to the reactant valley and $\kappa(T)$ is the tunnelling coefficient that accounts for reactive trajectories that do not reach the classical threshold energy. Because the definition of the selected reaction coordinate (z) involves only coordinates of atoms from the reacting subsystem (substrate and cofactor), the dynamic impact of protein motions in the chemical step will be reflected in a recrossing coefficient that deviates from unity. The two terms of eq 2 can be obtained from QM/MM simulations (for details, see the Supporting Information). The same approach has previously been used to analyze the reactivity of wild type EcDHFR.[28]

The calculated rate constants for wild type EcDHFR (Table 2) are in excellent agreement with the experimental data,[28] while the values calculated for EcDHFR-N23PP/S148A are somewhat smaller than those measured by experiment. This is most probably due to a small systematic error in the energy function that leads to an overestimation of the free energy barrier. However, the difference is below 1 kcal·mol^{-1} and of the same order of magnitude as the statistical uncertainty associated with the free energies (Table 2). The calculations show clearly that the reduction in the rate constant in EcDHFR-N23PP/S148A relative to the wild type enzyme is essentially due to the increase in the quasi-classical activation free energy (Table 2), while the recrossing transmission coefficients are similar but distinct and tunnelling contributions are also very close in both wild type and variant enzyme (Table 2). These results support the previous conclusions that the reduction in the rate constant in EcDHFR-N23PP/S148A is not due to the impact of a "dynamic knockout" on the chemical step but to a change in the properties of the equilibrium in the reactant state (RS) and transition state (TS) ensembles.[27,37] As discussed below, mutations provoke subtle changes in average protein−protein and protein−substrate/cofactor interactions that affect the free energy barrier.

Comparison of heavy and light EcDHFR and EcDHFR-N23PP/S148A shows that the difference in phenomenological rate constants between isotopomers arises from differences in the recrossing coefficients (γ). Tunnelling contributions (κ) are not affected by the change in mass between the light and heavy enzymes, which is not in agreement with proposals of tunnelling enhancement by dynamic coupling.[2−4] According to our simulations, such compressive "promoting" motions need not be invoked to explain the observed changes in reactivity; this finding is in agreement with recent work that revealed that "promoting vibrations" do not drive hydride transfer in the active site of EcDHFR.[38] However, the value of the recrossing transmission coefficient (γ) reflects the subtle coupling of protein environmental motions to the reaction coordinate in a way that is only apparent via a global description of all atomic positions. When the mass of the enzyme is increased, these motions are slower and the chemical system is not so efficient in relaxing to the reactant or product valleys after crossing the TS. This leads to an increase in the number of recrossings and therefore a reduction of the value of γ in the heavy enzyme relative to its light counterpart.

Simulations successfully reproduce the larger enzyme isotope effects obtained experimentally for the variant (Table 2). The computed k^{LE}/k^{HE} values for the variant and wild type enzymes are 1.26 and 1.16, while the experimental values are 1.37 and 1.10. The increased enzyme KIE for EcDHFR-N23PP/S148A relative to the wild type provides strong support for enhanced coupling of protein environmental motions to the reaction coordinate in the variant so that the chemical step becomes more sensitive to global protein motions in the modified enzyme. Our results show that the dynamic impact of protein motions on the chemical step is in fact larger in EcDHFR-

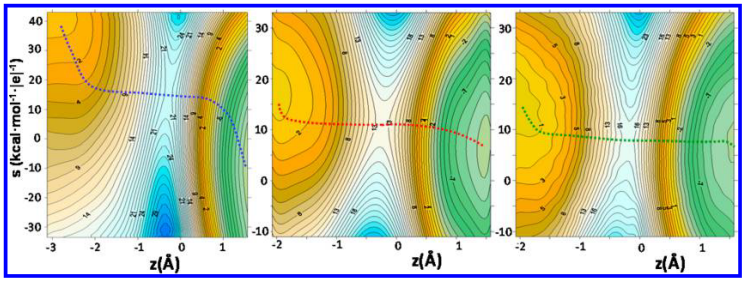


Figure 3. Free energy surfaces corresponding to the hydride transfer from NADPH to protonated H_2F in aqueous solution (left), wild type EcDHFR (middle), and EcDHFR-N23PP/S148A (right). See text for a full description of the two coordinates. Each isoenergetic line represents a 1 kcal·mol^{-1} increase in free energy. The dotted lines represent the minimum free energy paths on the free energy surfaces obtained from the gradient of the surface.

N23PP/S148A than in the wild type enzyme, as reflected by the larger deviation from unity observed in the recrossing transmission coefficients of 0.53 and 0.42 for light and heavy EcDHFR-N23PP/S148A versus 0.57 and 0.49 for light and heavy EcDHFR. Further simulations were conducted to rationalize this unexpected behavior of the "dynamic knockout" variant.

Protein Motions in EcDHFR and EcDHFR-N23PP/S148A. To gain insight into the role of protein motions in EcDHFR and EcDHFR-N23PP/S148A, simulations of the free energy landscape of the enzymes were performed. Free energy surfaces (FESs) for hydride transfer in wild type EcDHFR and EcDHFR-N23PP/S148A and in the absence of an enzyme in aqueous solution were traced as a function of two coordinates (Figure 3; see the Supporting Information for details), namely, the solute (or chemical) coordinate (z) that corresponds to the transfer of the hydride and a solvent (or environmental) coordinate (s) that captures those environmental motions relevant for the chemical process.[49] This coordinate is obtained as the antisymmetric combination of the electrostatic potential created by the protein and aqueous environment on the donor and acceptor carbon atoms ($s = V(C4_{NADPH}) - V(C6_{H2F})$) and is analogous to the solvent polarization in Marcus theory of electron transfer reactions.[50]

A more complete description of the chemical process can be obtained by following the minimum free energy paths traced on the FESs (Figure 3). In all cases, the reaction starts with a change in the solvent coordinate, then essentially moves along the chemical coordinate to pass through the TS, and finishes with a new change in the solvent coordinate, leading to the relaxed products. Thus, the reaction proceeds with participation of the solvent coordinate, mostly before and after crossing the TS region. This explains the small deviation from unity observed in all the recrossing transmission coefficients reported in the previous section.

Environmental motions involved in the evolution from the Michaelis complex to the TS can be characterized by means of the frequency associated with the solvent coordinate. To this end, the force constants were evaluated from a parabolic fit of the free energy landscape along the solvent coordinate and the associated effective masses were derived from the equipartition theorem. From these data, the resulting frequencies for the three environments can be calculated. The values of the force constants found for the enzymes (2.4×10^4 and 2.3×10^4 kcal^{-1}·mol·e^2 for EcDHFR and EcDHFR-N23PP/S148A, respectively) are significantly larger than those for aqueous solution (2.6×10^3 kcal^{-1}·mol·e^2). This is to be expected, as deforming the environment is energetically more demanding in enzymes than in solution due to the existence of a covalent structure in the proteins that does not exist in solution. As observed, the mass associated with the solvent coordinate is also significantly larger in the case of the two enzymes (1.2×10^{31} and 1.3×10^{31} kcal^{-1}·mol·e^2·s^2 for EcDHFR and EcDHFR-N23PP/S148A, respectively) than in solution (2.6×10^{30} kcal^{-1}·mol·e^2·s^2). The combined effect of the larger force constants and the larger associated mass for the enzymes means that the frequencies associated with the environmental motions are similar in all scenarios, and practically identical in EcDHFR and EcDHFR-N23PP/S148A (240 cm^{-1} in EcDHFR, 230 cm^{-1} in EcDHFR-N23PP/S148A, and 170 cm^{-1} in solution). Thus, environmental motions relevant for the hydride transfer have similar time scales in the three cases (of the order of picoseconds or faster). Analysis of these frequency modes reveals them to be associated with hydrogen bond pattern rearrangement around the chemical system (substrate and cofactor). The characteristic frequencies obtained for the solvent coordinate in the variant DHFR are not consistent with a mutation-induced change of the protein dynamics that could have noticeable consequences for the rate of hydride transfer. Femtosecond−picosecond time scale motions are the motions that can couple to the chemical step. This finding does however not exclude an effect from the mutations on the millisecond protein dynamics,[36] but such changes of the dynamics do not affect crossing of the barrier to hydrogen transfer.

The decrease in the rate constant observed in EcDHFR-N23PP/S148A relative to the wild type enzyme can be attributed almost exclusively to an increase in the quasi-classical activation free energy of the chemical step (Table 2). Introducing mutations into the enzyme can provoke changes in the RS and/or TS ensembles and thus in the free energy differences between them. An analysis of averaged geometries has been carried out from 2 ns QM/MM MD simulations at the TS and RS of EcDHFR and EcDHFR-N23PP/S148A. In the case of EcDHFR-N23PP/S148A, the mutations alter some protein−protein and protein−cofactor interactions established by residues belonging to the M20 loop at the reaction TS (Table S5, Supporting Information). In particular, the averaged Glu17NH···OδAsp122 distance at the TS of EcDHFR-N23PP/

S148A is 2.4 ± 0.6 Å, while this hydrogen bond interaction between the M20 and FG loops is significantly weaker in the TS of wild type EcDHFR (the averaged distance is 4.0 ± 0.7 Å). The fact that this hydrogen bond is strongly formed only in the TS of EcDHFR-N23PP/S148A could contribute to the smaller enthalpic and larger entropic barrier found experimentally for the variant.[27]

Other changes in the interactions established by the M20 loop with the cofactor or the substrate are observed upon mutation. Met20 has a reduced capability to form a S···HN hydrogen bond with the amide group of the cofactor at the TS in the variant. The averaged distances are 2.6 ± 0.3 and 3.0 ± 0.4 Å in the TSs of EcDHFR and EcDHFR-N23PP/S148A, respectively (Table S5, Supporting Information). From an electronic point of view, the formation of this hydrogen bond favors hydride transfer from the cofactor to the substrate.[51] Met20 has also been proposed to play an important role at the cofactor/substrate interface.[32] Dihedral angles of the Met20 side chain (Figure S7 and Table S5, Supporting Information) show the preference for different conformations of these residues in the wild type and variant enzymes. In support of these findings, differences in the positioning of the sulfur atom of Met20 and in the preferred conformation of its side chain between wild type EcDHFR and EcDHFR-N23PP/S148A have been observed experimentally.[36] Finally, other residues of the M20 loop such as Asn18 and Ala19 are farther from the cofactor at the TS of the variant than in wild type EcDHFR (Table S5, Supporting Information), confirming the disruption of stabilizing interactions established by this loop in the TS by mutating the enzyme.

Differences in the equilibrium fluctuations of the protein residues between wild type and variant were analyzed from the 2 ns QM/MM MD simulations of the TS by means of the root-mean-squared fluctuation (RMSF) of all residues of the proteins. The differences observed between the RMSF of residues in EcDHFR and EcDHFR-N23PP/S148A (Figure 4) are never larger than 0.5 Å in absolute value. Thus, both proteins show similar flexibility on the time scale relevant to catalysis (*vide supra*) including in the flexible *β*FG, *β*GH, and M20 loops (Figure 4A). This result is in agreement with a recent evolutionary study of DHFRs from different species, which showed that mutations do not cause large changes to the equilibrium fluctuations but that subtle changes to the equilibrium conformational sampling alter the free energy barrier of the enzymatic reaction,[52] as reflected also in our results (Table 2 and Table S5, Supporting Information). Our simulations demonstrate that residues in the active site are slightly more mobile for EcDHFR-N23PP/S148A than for the wild type enzyme in the TS (Figure 4B), reflecting the fact that protein–substrate interactions established in the active site at the TS are not as well optimized in the variant. This is consistent with a recent computational analysis of X-ray crystallographic results, which shows increased conformational heterogeneity in the active site of EcDHFR-N23PP/S148A compared to the wild type enzyme.[53]

While differences between the rate constants of EcDHFR and EcDHFR-N23PP/S148A are mostly due to changes in the equilibrium properties of the enzymes, the fact that the active site of the variant is less well set up for catalysis than the active site of the wild type also has an impact on protein motions coupled to the reaction, as previously deduced from the comparison of the recrossing transmission coefficients (Table 2). The connection between mutations and dynamics can be rationalized observing the behavior of the environmental coordinate in the FESs (Figure 3). The ordering of the barriers without vibrational corrections is 14.6, 17.5, and 23 kcal·mol^{-1} for EcDHFR, EcDHFR-N23PP/S148A, and the reaction in aqueous solution, respectively. These values can be correlated with the displacement along the solvent coordinate s needed to reach the TS from the RS ($\Delta s^{\ddagger} = s(\mathrm{TS}) - s(\mathrm{RS})$). The $\Delta s^{\ddagger}$ values obtained from the FESs for the reaction in EcDHFR, EcDHFR-N23PP/S148A, and in solution are 5, 7, and 27 kcal·mol^{-1}·e^{-1}, respectively. While the two enzymes stabilize the RS at a solvent coordinate closer to the value needed to reach the TS, in aqueous solution the reactants are found at a much larger value of s (∼40 kcal·mol^{-1}·e^{-1}). That is, at the RS, both enzymes already provide an environment much more conducive to hydride transfer than in aqueous solution. In other words, the enzymatic active sites are preorganized to favor hydride transfer. Interestingly, the reorganization needed to reach the TS is larger for the variant than the wild type, as reflected in the larger value of $\Delta s^{\ddagger}$. Hence, the participation of protein motions in the reaction coordinate in EcDHFR-N23PP/S148A is larger than that in wild type EcDHFR. This observation agrees with the relative values of the recrossing transmission coefficients (Table 2) for the variant and the wild type that reflect a larger impact of protein motions in the former. Thus, mutations that reduce the catalytic properties of the enzymatic active site increase the participation of protein motions in the chemical step because the new active site requires a larger rearrangement to reach the TS.

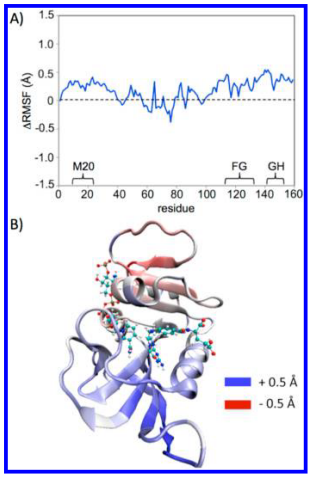


Figure 4. Differences of the femtosecond–picosecond root-mean-square fluctuations (RMSFs) between EcDHFR-N23PP/S148A and EcDHFR calculated for the backbone Ca atoms in the TSs. (A) RMSF difference versus residue number. (B) Projection on the protein backbone using a color scale: red represents regions of the protein that are more mobile in the EcDHFR than in EcDHFR-N23PP/S148A, while blue represents the opposite behavior.

CONCLUSIONS

The present investigation has used a combination of experimental and computational techniques to analyze the dynamic contributions to catalysis in EcDHFR-N23PP/S148A by comparing the heavy (^{15}N, ^{13}C, ^{2}H isotopically substituted) enzyme with its light (natural isotopic abundance) counterpart. The good agreement between the experimental and computationally derived results allows a deeper insight into the role of protein motions in enzyme catalysis. Interestingly, the impact of protein motions on catalysis is larger in EcDHFR-N23PP/S148A, previously called a "dynamic knockout" enzyme,[36] than in the wild type, as reflected in the smaller value of the recrossing transmission coefficient and the increased value of k^{LE}/k^{HE} for EcDHFR-N23PP/S148A relative to EcDHFR. Simulations that explicitly consider the changes in the environment during the chemical step show that the variant enzyme is less well set up to accommodate the chemical reaction and thus a higher degree of reorganization is required to go from the RS to the TS. This effect provokes a larger participation of protein motions in the reaction coordinate, as reflected in the smaller value of the transmission coefficient. Although millisecond conformational fluctuations in the active site of EcDHFR are lost in the N23PP/S148A variant,[36] these motions are not directly relevant to hydride transfer, as demonstrated by our characterization of the solvent coordinate changes during the chemical step. In fact, at the TS, the mobility of active site residues on the fs−ps time scale is slightly larger in EcDHFR-N23PP/S148A than in the wild type. Unlike in other enzymes, where motions on one time scale may promote those on another,[14,15] the N23PP/S148A mutation has very different effects on protein motions depending on the time scale. Motions on the μs−ms time scale are lost (at least in the M20 loop),[36] those on the ps−ns time scale are unaffected,[36] while those on the fs−ps time scale are enhanced by the mutation. EcDHFR-N23PP/S148A is therefore only a dynamic knockout on the time scale typical for conformational motions, but it is a "dynamic knock-in" at the level of the chemical step. The increased dynamic coupling of enzyme motions to the chemical coordinate is in fact detrimental to catalysis. Our findings most likely also apply to other enzyme-catalyzed reactions, although further studies are required to identify the nature of the motions that contribute to the recrossing coefficient in each case before predictive general conclusions can be drawn.

ASSOCIATED CONTENT

Supporting Information

Experimental methods; mass spectra of purified proteins; circular dichroism spectra; tabulated experimental data for k_H, k_{cat}, K_m, and enzyme KIEs; pH dependence of k_H; computational methods; and additional computational results. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

ignacio.tunon@uv.es
moliner@uji.es
allemannrk@cf.ac.uk

Author Contributions

[‖]J.J.R.-P. and L.Y.P.L.: These authors contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by grant BB/J005266/1 (R.K.A.) from the UK Biotechnology and Biological Sciences Research Council (BBSRC), by the Vice Chancellor Fund of Cardiff University, by the Spanish Ministerio de Economía y Competitividad (project CTQ2012-36253-C03), by MICINN (project CTQ2009-14541-C02), by Generalitat Valenciana (projects ACOMP/2012/119, ACOMP/2012/243, GV/2012/044, and Prometeo/2009/053), and by Universitat Jaume I-Bancaixa (projects P1·1A2010-08 and P1·1B2011-23). J.J.R.-P. thanks the Spanish Ministerio de Ciencia e Innovación for a "Juan de la Cierva" contract. R.G.-M. acknowledges a FPU fellowship of the Ministerio de Economía y Competitividad. We acknowledge Prof. D. G. Truhlar for helpful comments on the original manuscript. I.T. acknowledges many helpful discussions with Prof. J. T. Hynes and D. Laage during his sabbatical stay at the Ecole Normale Supérieure (Paris). We acknowledge the computational facilities of Universitat Jaume I, Universitat de València (Tirant Supercomputer), and the Spanish Superconducting Network (Picasso Supercomputer).

REFERENCES

(1) Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 444−448.
(2) Nagel, Z. D.; Klinman, J. P. *Nat. Chem. Biol.* **2009**, *5*, 543−550.
(3) Antoniou, D.; Caratzoulas, S.; Kalyanaraman, C.; Mincer, J. S.; Schwartz, S. D. *Eur. J. Biochem.* **2002**, *269*, 3103−3112.
(4) Scrutton, N. S.; Basran, J.; Sutcliffe, M. J. *Eur. J. Biochem.* **1999**, *264*, 666−671.
(5) Allemann, R. K.; Evans, R. M.; Loveridge, E. J. *Biochem. Soc. Trans.* **2009**, *37*, 349−353.
(6) Sikorski, R. S.; Wang, L.; Markham, K. A.; Rajagopalan, P. T. R.; Benkovic, S. J.; Kohen, A. *J. Am. Chem. Soc.* **2004**, *126*, 4778−4779.
(7) Wang, L.; Tharp, S.; Selzer, T.; Benkovic, S. J.; Kohen, A. *Biochemistry* **2006**, *45*, 1383−1392.
(8) Anandarajah, K.; Schowen, K. B.; Schowen, R. L. *Z. Phys. Chem.* **2008**, *222*, 1333−1347.
(9) Hay, S.; Pudney, C. R.; Scrutton, N. S. *FEBS J.* **2009**, *276*, 3930−3941.
(10) Chowdhury, S.; Banerjee, R. *J. Am. Chem. Soc.* **2000**, *122*, 5417−5418.
(11) Heyes, D. J.; Sakuma, M.; de Visser, S. P.; Scrutton, N. S. *J. Biol. Chem.* **2009**, *284*, 3762−3767.
(12) Henzler-Wildman, K.; Kern, D. *Nature* **2007**, *450*, 964−972.
(13) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. *Science* **1991**, *254*, 1598−1603.
(14) Henzler-Wildman, K. A.; Lei, M.; Thai, V.; Kerns, S. J.; Karplus, M.; Kern, D. *Nature* **2007**, *450*, 913−916.
(15) Pineda, J. R. E. T.; Antoniou, D.; Schwartz, S. D. *J. Phys. Chem. B* **2010**, *114*, 15985−15990.
(16) Liu, H. B.; Warshel, A. *Biochemistry* **2007**, *46*, 6011−6025.
(17) Pisliakov, A. V; Cao, J.; Kamerlin, S. C. L.; Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 17359−17364.
(18) Glowacki, D. R.; Harvey, J. N.; Mulholland, A. J. *Nat. Chem.* **2012**, *4*, 169−176.
(19) Roca, M.; Oliva, M.; Castillo, R.; Moliner, V.; Tunon, I. *Chem.—Eur. J.* **2010**, *16*, 11399−11411.
(20) Ruiz-Pernia, J. J.; Tunon, I.; Moliner, V.; Hynes, J. T.; Roca, M. *J. Am. Chem. Soc.* **2008**, *130*, 7477−7488.
(21) Doll, K. M.; Finke, R. G. *Inorg. Chem.* **2003**, *42*, 4849−4856.
(22) Doll, K. M.; Bender, B. R.; Finke, R. G. *J. Am. Chem. Soc.* **2003**, *125*, 10877−10884.

(23) Loveridge, E. J.; Allemann, R. K. *Biochemistry* **2010**, *49*, 5390−5396.
(24) Loveridge, E. J.; Tey, L. H.; Allemann, R. K. *J. Am. Chem. Soc.* **2010**, *132*, 1137−1143.
(25) Loveridge, E. J.; Allemann, R. K. *ChemBioChem* **2011**, *12*, 1258−1262.
(26) Loveridge, E. J.; Tey, L.-H.; Behiry, E. M.; Dawson, W. M.; Evans, R. M.; Whittaker, S. B.-M.; Guenther, U. L.; Williams, C.; Crump, M. P.; Allemann, R. K. *J. Am. Chem. Soc.* **2011**, *133*, 20561−20570.
(27) Loveridge, E. J.; Behiry, E. M.; Guo, J.; Allemann, R. K. *Nat. Chem.* **2012**, *4*, 292−297.
(28) Luk, L. Y. P.; Ruiz-Pernia, J. J.; Dawson, W. M.; Roca, M.; Loveridge, E. J.; Glowacki, D. R.; Harvey, J. N.; Mulholland, A. J.; Tuñón, I.; Moliner, V.; Allemann, R. K. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 16344−16349.
(29) Benkovic, S. J.; Hammes-Schiffer, S. *Science* **2003**, *301* (80), 1196−1202.
(30) Nashine, V. C.; Hammes-Schiffer, S.; Benkovic, S. J. *Curr. Opin. Chem. Biol.* **2010**, *14*, 644−651.
(31) Liu, H. B.; Warshel, A. *J. Phys. Chem. B* **2007**, *111*, 7852−7861.
(32) Khavrutskii, I. V; Price, D. J.; Lee, J.; Brooks, C. L. *Protein Sci.* **2007**, *16*, 1087−1100.
(33) Boehr, D. D.; McElheny, D.; Dyson, H. J.; Wright, P. E. *Science* **2006**, *313*, 1638−1642.
(34) Sawaya, M. R.; Kraut, J. *Biochemistry* **1997**, *36*, 586−603.
(35) Fierke, C. A.; Johnson, K. A.; Benkovic, S. J. *Biochemistry* **1987**, *26*, 4085−4092.
(36) Bhabha, G.; Lee, J.; Ekiert, D. C.; Gam, J.; Wilson, I. A.; Dyson, H. J.; Benkovic, S. J.; Wright, P. E. *Science* **2011**, *332*, 234−238.
(37) Adamczyk, A. J.; Cao, J.; Kamerlin, S. C. L.; Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 14115−14120.
(38) Dametto, M.; Antoniou, D.; Schwartz, S. D. *Mol. Phys.* **2012**, *110*, 531−536.
(39) Born, M.; Oppenheimer, J. R. *Ann. Phys.* **1927**, *389*, 457−484.
(40) Kipp, D. R.; Silva, R. G.; Schramm, V. L. *J. Am. Chem. Soc.* **2011**, *133*, 19358−19361.
(41) Silva, R. G.; Murkin, A. S.; Schramm, V. L. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 18661−18665.
(42) Pudney, C. R.; Guerriero, A.; Baxter, N. J.; Johannissen, L. O.; Waltho, J. P.; Hay, S.; Scrutton, N. S. *J. Am. Chem. Soc.* **2013**, *135*, 2512−2517.
(43) Toney, M. D.; Castro, J. N.; Addington, T. A. *J. Am. Chem. Soc.* **2013**, *135*, 2509−2511.
(44) Venkitakrishnan, R. P.; Zaborowski, E.; McElheny, D.; Benkovic, S. J.; Dyson, H. J.; Wright, P. E. *Biochemistry* **2005**, *44*, 5948.
(45) Glasstone, S.; Laidler, K. J.; Eyring, H. *The theory of rate processes: the kinetics of chemical reactions, viscosity, diffusion and electrochemical phenomena*; McGraw-Hill: New York, 1941.
(46) Keck, J. C. *Adv. Chem. Phys.* **1967**, *13*, 85−121.
(47) Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J. *J. Phys. Chem.* **1996**, *100*, 12771−12800.
(48) Alhambra, C.; Corchado, J.; Sanchez, M. L.; Garcia-Viloca, M.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. B* **2001**, *105*, 11326−11340.
(49) García-Meseguer, R.; Martí, S.; Ruiz-Pernía, J. J.; Moliner, V.; Tuñón, I. *Nat Chem* **2013**, *5*, 566−571.
(50) Marcus, R. A. *Annu. Rev. Phys. Chem.* **1964**, *15*, 155−196.
(51) Garcia-Viloca, M.; Truhlar, D. G.; Gao, J. *Biochemistry* **2003**, *42*, 13558−13575.
(52) Liu, C. T.; Hanoian, P.; French, J. B.; Pringle, T. H.; Hammes-Schiffer, S.; Benkovic, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 10159−10164.
(53) Van den Bedem, H.; Bhabha, G.; Yang, K.; Wright, P. E.; Fraser, J. S. *Nat. Methods* **2013**, *10*, 896−902.

THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

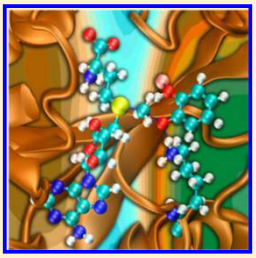
Linking Electrostatic Effects and Protein Motions in Enzymatic Catalysis. A Theoretical Analysis of Catechol *O*-Methyltransferase

Rafael García-Meseguer,[†] Kirill Zinovjev,[†] Maite Roca,[†] Javier J. Ruiz-Pernía,*[,‡] and Iñaki Tuñón*[,†]

[†]Departament de Química Física, Universitat de València, 46100 Burjassot, Spain

[‡]Departament de Química Física i Analítica, Universitat Jaume I, 12071 Castelló, Spain

ABSTRACT: The role of protein motions in enzymatic catalysis is the subject of a hot scientific debate. We here propose the use of an explicit solvent coordinate to analyze the impact of environmental motions during the reaction process. The example analyzed here is the reaction catalyzed by catechol *O*-methyltransferase, a methyl transfer reaction from *S*-adenosylmethionine (SAM) to the nucleophilic oxygen atom of catecholate. This reaction proceeds from a charged reactant to a neutral product, and then a large electrostatic coupling with the environment could be expected. By means of a two-dimensional free energy surface, we show that a large fraction of the environmental motions needed to attain the transition state happens during the first stages of the reaction because most of the environmental motions are slower than changes in the substrate. The incorporation of the solvent coordinate in the definition of the transition state improves the transmission coefficient and the committor histogram in solution, while the changes are much less significant in the enzyme. The equilibrium solvation approach seems then to work better in the enzyme than in aqueous solution because the enzyme provides a preorganized environment where the reaction takes place.



1. INTRODUCTION

Prof. W. L. Jorgensen has pioneered the development and application of computational methodologies for the calculation of free energy changes associated with chemical transformations taking place in condensed phases. Molecular simulations were used to calculate free energy differences related to solvation,[1] ion-pairing,[2] acid−base[3] and conformational equilibria,[4] and protein−ligand binding.[5,6] Prof. Jorgensen also pioneered the use of free energy profiles to rationalize environmental effects on the rate of chemical reactions with seminal studies on the S_N2 process in solution.[7,8] According to these studies, the rate constant diminution observed for the identity $Cl^- + CH_3Cl$ reaction in aqueous solution relative to the gas-phase can be explained as the consequence of an increase in the activation free energy due to the weaker solvation interactions established at the transition state (TS). This interpretation relies on the validity of the equilibrium solvation assumption to describe the TS of the chemical reaction, a hypothesis frequently invoked in applications of transition state theory (TST).[9,10] In this regard, a correct definition of the TS, or equivalently an adequate selection of the reaction coordinate, can be crucial to ensure the correct use of TST to determine the rate constant of a chemical reaction. In their study, Jorgensen and co-workers employed the antisymmetric combination of the distances of the broken and formed bonds as the reaction coordinate, assuming then that all the solvent degrees of freedom equilibrate at each value of this coordinate. Solvent dynamics is composed of a mixture of different kind of motions (including hindered translations and rotations), and not all of them are fast enough to be considered in equilibrium with the antisymmetric stretching reaction coordinate. The influence on nonequilibrium solvation effects on the reaction rate for the S_N2 reaction in water was addressed some years later by Hynes and co-workers[11,12] and Warshel and co-workers[13] who found that the fate of a trajectory is largely dependent on the solvent configuration at the TS. This effect can be incorporated into TST either by the calculation of the transmission coefficient (κ) associated with the reaction coordinate or by changing to a new reaction coordinate incorporating not only solute degrees of freedom but also information about the solvent. Although the transmission coefficient corrections calculated for the S_N2 reaction in water ($\kappa \approx 0.5$) only have a modest quantitative effect on the reaction rate constant, these studies show that solvent participation in the reaction process may be different from that obtained from the equilibrium picture, especially for reaction systems strongly coupled to the surroundings.[11]

The role of environmental motions during the chemical reaction is at the heart of the debate in the field of enzymatic chemistry. It is a well-accepted fact that proteins are not rigid scaffolds where catalysis takes place but that they must be flexible enough to permit the evolution among different conformations along the catalytic cycle.[14,15] However, the question of how molecular motions within the protein structure may influence the rate constant of the chemical step has been the subject of many contradictory and sometimes controversial studies during the past years.[16−36] Even with acceptance that the enzymatic active site is preorganized to optimize the

Special Issue: William L. Jorgensen Festschrift

Received: June 10, 2014
Revised: August 26, 2014
Published: August 27, 2014

electrostatic interaction with the TS,[37] changes during the chemical step do not merely affect the chemical system but also the protein. The protein environment is reorganized during the evolution of the system from the reactant state (RS) to the TS,[26] which implies the participation of protein motions in the chemical transformation.[21,26,38,39] Different kinds of protein motions participating in the chemical transformation have been identified from molecular simulations[27,30,39−41] or invoked to explain experimental observations such as kinetic isotope effects[32] and consequences of mutations on the rate constant.[29] These motions include both vibrational motions,[32,33] taking place on a time scale similar to the barrier crossing event, and very slow conformational changes[29,42] that could be considered as frozen during barrier crossing.

The main question to solve with respect to the role of protein motions in enzymatic catalysis is to determine if the equilibrium treatment of protein motions provides a reasonable description for enzymatic reactions[34] or if an explicit dynamical treatment is required for at least some of these motions.[33,43] It has been recently claimed that the equilibrium assumption inherent to traditional TST can result in TS ensembles significantly different from those obtained by means of transition path sampling techniques, where this assumption is not made.[36] Obviously the validity of the equilibrium approach is intimately linked to the nature of the reaction coordinate chosen to represent the advance of the system along the chemical process. Empirical valence bond (EVB) simulations[44] typically use the energy gap coordinate, where both solute/substrate and solvent/protein motions are incorporated. On the other hand, quantum mechanics/molecular mechanics (QM/MM) simulations that employ semiempirical, density functional, or ab initio Hamiltonians are usually based on the selection of a reaction coordinate defined in terms of geometrical valence coordinates of the solute/substrate, such as bond distances or combinations of them. These last coordinates are usually faster than some environmental motions that could be active during the reaction progress. Even so, values of transmission coefficients of enzymatic reactions calculated for these kinds of reaction coordinates usually range between 0.5 and 0.9,[25,45] indicating a minor participation of protein motions during barrier crossing. However, these findings do not discard a more active participation of the environment in other stages of the chemical process.

In order to address the problem of the participation of environmental motions in enzymatic catalysis, we proposed to explore the reaction using a reaction coordinate ($\xi(q,s)$) based on a combination of solute (q) and solvent (s) degrees of freedom.[35] For the solute coordinate a simple valence coordinate can be used, while for the solvent we propose a collective variable expressed as a function of the electrostatic potential created by the environment on key atoms of the chemical system.[46] The analysis of the free energy surfaces (FESs) traced as a function of the two coordinates shows that the reaction progress must be expressed as a combination of both coordinates and that the equilibrium approach fails to reproduce the minimum free energy path found on this surface.[35] However, for the enzymatic examples analyzed so far, barrier crossing takes place essentially along the solute coordinate and the equilibrium picture provides a reasonable estimation of the TS ensemble and thus also of the activation free energy and the rate constant.[35] In fact, deviations from this equilibrium approach are found to be larger for reaction in aqueous solution because a larger reorganization of the environment is required to reach the TS.[35,46]

In this work we analyze the performance of the electrostatic potential coordinate to study the role of environmental motions in the case of very strong electrostatic solute−solvent coupling. We present a theoretical analysis of the methyl transfer reaction from *S*-adenosylmethionine (SAM) to the nucleophilic hydroxylate oxygen atom of catechol catalyzed by catechol *O*-methyltransferase (COMT, EC 2.1.1.6).[47] This reaction (see Figure 1) is a direct bimolecular S_N2 process that can be

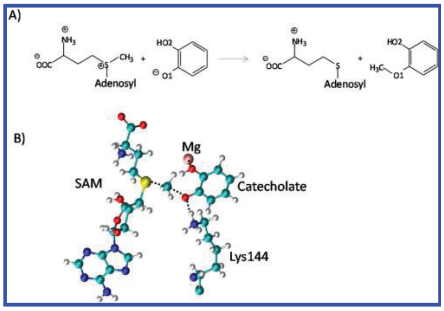


Figure 1. (A) Schematic representation of the methylation reaction of catecholate by SAM. (B) Transition state structure obtained for the enzymatic reaction showing the interactions of the substrate with the Mg^{2+} ion and with Lys144.

formally considered as an inverse Menshutkin reaction where ionic reactants proceed to neutral products,[48−54] and then important reorganization effects could take place in the surroundings of the chemical system. Previous theoretical studies on this system showed that the enzymatic active site provides an electrostatic environment more adequate for the reaction than in aqueous solution,[41,55] where solute−solvent interactions provide TS destabilization with respect to charged reactant state.[56] Calculation of the transmission coefficient and the friction kernel associated with a simple valence coordinate for the methyl transfer showed a stronger coupling of solvent motions at the TS in solution than in the enzymatic environment.[39] However, these analyses did not explicitly address the role of protein motions during the whole reaction process. The experimental observation of inverse secondary α-deuterium kinetic isotope effects (KIEs)[48,57] in rat liver COMT and the effects of site-specific active site mutations on the rate constant and KIEs in human COMT[32] were interpreted as the signature of the participation of environmental motions in the reaction that could favor the compression of the methyl donor−acceptor distance in the enzyme with respect to the in-solution process. However, QM/MM simulations were able to reproduce the inverse KIEs without observing any reduction in the donor−acceptor distance at the enzymatic TS determined within the equilibrium solvation approximation.[58] Simulations presented in this paper allow a deeper discussion on the role of environmental motions for a reaction where the energy strongly depends on coupled solute−solvent terms (strong coupling), both in solution and in the enzyme. Analyses of the FESs obtained as a function of the electrostatic (solvent) and valence (solute) coordinates show the participation of protein motions

dx.doi.org/10.1021/jp505746x | *J. Phys. Chem. B* 2015, 119, 873−882

at different stages of the reaction advance. In general, environmental motions are required to prepare a good electrostatic environment where the reaction can proceed. These motions precede solute motions, and barrier crossing takes place mostly along the valence coordinate, especially in the enzymatic reaction. The equilibrium-solvation TS ensemble characterized along the solute coordinate is thus a good approximation to a more rigorous TS ensemble characterized by explicit consideration of other (environmental) degrees of freedom.

2. METHODOLOGY

The initial structure for the simulations carried out in this work was taken from our previous studies on rat liver catechol *O*-methyltransferase and the counterpart process in aqueous solution.[39,41,55,58,59] Briefly, the system was prepared starting from the X-ray structure with Protein Data Bank code 1VID, which contains the ternary complex of the enzyme with the *S*-adenosylmethionine (SAM) cofactor and an inhibitor 3,5-dinitrocatechol.[60] Both nitro groups were removed to prepare the natural substrate which was ionized by proton transfer to neutral Lys144.[52] The whole system was placed inside a pre-equilibrated cubic box of water molecules of side 79.5 Å, deleting all water molecules with oxygen atoms found at 2.8 Å from any non-hydrogen atom of the protein. Two sodium ions where added to the enzymatic system to keep it electroneutral. To study the reaction in solution, we followed the same procedure placing the whole cofactor SAM and the substrate in a prequilibrated water box of side 55.8 Å.

We used a QM/MM computational scheme in which *S*-adenosylmethionine (SAM) and catecholate (63 atoms) are described using the AM1 semiempirical Hamiltonian.[61] This QM level overestimates the free energy barrier of the reaction but provides reasonable geometries for the stationary structures, with a slight underestimation of the distance between the carbon atom of the transferred methyl group to the nucleophilic oxygen atom of cathecholate (O1 in Figure 1).[59,62] The rest of the system forms the MM subsystem described by means of the AA-OPLS force field for the enzyme[63] and a modified version of the TIP3P potential for water molecules.[64,65] Except when indicated, all the simulations were carried out in the NVT ensemble using the Langevin–Verlet integrator with a time step of 1 fs and a target temperature of 300 K. A group-based switched cutoff radius between 12 and 14 Å was employed for all MM and QM-MM interactions. While the treatment of long-range electrostatic QM/MM interactions could be improved using Ewald summation,[66] we found that the use of an electrostatic cutoff was enough to reproduce the catalytic effect; this is the lowering in the activation free energy when passing from aqueous solution to the enzyme.[55] Thus, this treatment is enough for a comparative analysis between both environments.

In this work we obtained FESs using two different coordinates, a solute coordinate (q) and a solvent coordinate(s). The FES can be expressed as

$$W(q, s) = C' - k_B T \ln \int [\rho(x^N)\, \delta(q(x^N) - q_0)\, \delta(s(x^N) - s_0)]\, \mathrm{d}x^N \quad (1)$$

where $\rho(x^N)$ is the probability density of finding the system at the configuration x^N, k_B the Boltzmann constant, and T the absolute temperature. In this case, the solute coordinate is the antisymmetric combination of the distances of the SAM sulfur atom and catecholate oxygen to the carbon atom of the transferred methyl group ($q = d(\mathrm{SC}) - d(\mathrm{O1C})$); see Figure 1. The solvent coordinate was the antisymetric combination of the electrostatic potential created by the environment on sulfur and oxygen atoms:

$$s = V_{O1}(x^N) - V_S(x^N) = \sum_{j=1}^{M} \frac{Q_j}{|x_j - x_{O1}|} - \sum_{j=1}^{M} \frac{Q_j}{|x_j - x_S|} \quad (2)$$

where the sums run on the M sites of the environment with point charges Q_j. This is a collective coordinate that involves all the environmental atoms with an electrostatic influence on the donor or acceptor atoms. It must be noticed that, by construction, the solute and the solvent coordinates are not orthogonal, as they involve the coordinates of some common atoms (the methyl donor and acceptor). However, we checked that at the TSs, these coordinates are in practice almost orthogonal. The averaged angles between them, determined from the gradients in the mass-weighted Cartesian space, were 90.3 ± 0.5 and 95.4 ± 0.4 degrees in aqueous solution and in the enzyme, respectively.

The FESs that correspond to the reaction in aqueous solution and in the enzyme were obtained using umbrella sampling,[67] applying harmonic biasing potentials to the solute and the solvent coordinates:

$$V_q = \frac{1}{2} K_q (q - q_0)^2; \quad V_s = \frac{1}{2} K_s (s - s_0)^2 \quad (3)$$

With these constraints the system preferentially explores the most probable configurations around the reference coordinate values q_0 and s_0. These values are changed to sample the range of values of interest. The joint probability distributions of the two coordinates were obtained by means of the weighted histogram analysis method (WHAM).[68] A total of 2840 simulation windows were employed to trace the FES in aqueous solution, while for the enzyme 1420 windows were needed. The difference comes from the fact that the reaction takes place spanning a wider range of values of the solvent coordinate in aqueous solution than in the enzyme, as shown below. Each simulation window consisted of 5 ps of equilibration and 45 ps of production. The force constants used to keep the system at the reference values of the solute and solvent coordinates were 2500 kJ·mol^{-1} Å^{-2} and 0.01 kJ^{-1}·mol·|e|^2, respectively. These values were shown to provide a good control of the coordinates.[46]

The TS ensembles for the reaction in solution (TS_w) and in the enzyme (TS_e) were characterized on a monodimensional surface traced as a function of the solute coordinate ($TS_w(q)$ and $TS_e(q)$) and on the two-dimensional surface, function of the solute and solvent coordinates ($TS_w(q,s)$ and $TS_e(q,s)$). The approximate positions of the TS ensembles were first estimated from the corresponding FESs. Then 100 ps long QM/MM simulations were performed with a biasing potential added to restrain the system around the saddle point. Assuming that the region of the FES accessible to such a simulation can be well described within the harmonic approximation, we calculated the precise position of the saddle point ($\boldsymbol{\xi}^{TS}$) on the FESs as[69]

$$\boldsymbol{\xi}^{TS} = \mathbf{H}^{-1}(\beta^{-1}\mathbf{C}^{-1}\cdot\boldsymbol{\xi}(\mathbf{R})^b - \mathbf{K}\cdot\boldsymbol{\xi}^0) \quad (4)$$

dx.doi.org/10.1021/jp505746x | J. Phys. Chem. B 2015, 119, 873–882

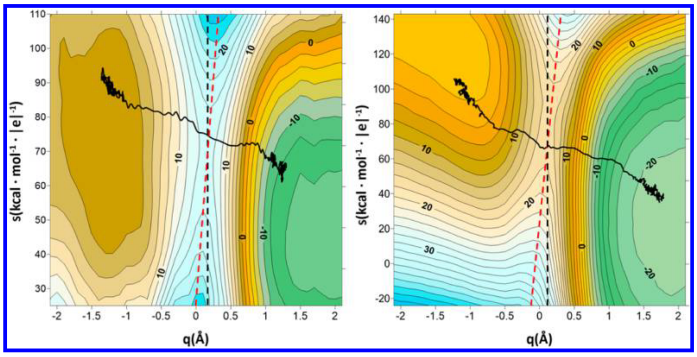


Figure 2. Free energy surfaces (isocontour lines are drawn each 2 kcal·mol^{-1}) corresponding to the enzymatic reaction (left) and in solution (right). Averaged reactive trajectories are shown as a continuous line. TS surfaces obtained using the solute or the solute and the solvent coordinates are indicated by dashed lines in black and red, respectively.

where $\beta = 1/(k_B T)$, $\mathbf{H}$ is the 2 × 2 Hessian matrix associated with the 2D-FES, $\mathbf{C}$ the covariance matrix, $\mathbf{K}$ the diagonal matrix of force constants, $\langle \boldsymbol{\xi}(\mathbf{R}) \rangle^b$ the average over the biased simulation, and $\boldsymbol{\xi}^0$ the reference position of the bias. The Hessian matrix is obtained as[69]

$$\mathbf{H} = \beta^{-1}\mathbf{C}^{-1} - \mathbf{K} \tag{5}$$

Once the Hessian matrices and TS positions are known, a biasing potential restraining the simulation to sample the TS ensemble can be constructed:

$$V^{TS}(\boldsymbol{\xi}(\mathbf{R})) = \frac{K^{TS}}{2}((\boldsymbol{\xi}(\mathbf{R}) - \boldsymbol{\xi}^{TS})\cdot\mathbf{v}_-)^2 \tag{6}$$

where K^{TS} is the force constant, which must be as large as possible to keep the system close to the TS, and $\mathbf{v}_-$ is the eigenvector of the Hessian that corresponds to the negative eigenvalue. With this bias, sets of 100 TS structures were obtained from MD simulations, saving those structures were $V^{TS} < 0.5k_B T$ and that were separated by at least 1 ps of simulation. From each TS structure 50 trajectories with random initial velocities taken from a Maxwell−Boltzmann distribution at T = 300 K were integrated both forward and backward in time in the NVE ensemble using the velocity Verlet integrator to obtain the probability to commit to the reactants and products basins.

3. RESULTS

Free Energy Surfaces. The AM1/MM FESs corresponding to the methyl transfer in aqueous solution and in the active site of COMT as a function of the solute (q) and solvent (s) coordinates are presented in Figure 2. The solute coordinate evolves from negative values at the reactant state (when the methyl group is bonded to the sulfur atom of the cofactor) to positive values at the products state (when the methyl group is bonded to the nucleophile oxygen atom of catecholate). The solvent coordinate changes from positive values at the reactant state to smaller values at the product state. The electrostatic potential created by the environment at the reactant state on the nucleophile atom of catecholate is positive, while the potential on the sulfur atom of SAM is negative as a consequence of the charge held by these atoms. The electrostatic potential on these two atoms diminishes at the product state where charge separation has been annihilated. Average reactive trajectories initiated at the TS have been projected on the FESs and represented with a continuous line. The positions of the TS ensembles on the two-dimensional surfaces (TS(q,s)) are represented by means of a black dashed line, while the TSs obtained in the monodimensional or equilibrium solvation description (TS(q)) correspond to the vertical red dashed line. It should be noticed that the average trajectory does not necessarily cross exactly at the saddle point on the FES because of the asymmetries of the free energy landscape in the directions orthogonal to the reaction advance.

Both FESs show a saddle point at very similar values of the solute and solvent coordinates. In the enzyme this saddle point is located at q_e^{TS} = 0.13 Å and s_e^{TS} = 67 kcal·mol^{-1}·|e|$^{-1}$, while in aqueous solutions the values found for the two coordinates are q_w^{TS} = 0.08 Å and s_w^{TS} = 78 kcal·mol^{-1}·|e|$^{-1}$. Larger differences appear at the reactant region. In aqueous solution the FES does not display a minimum at the reactant region, at least for the range of values explored for the solute coordinate. By use of a very similar computational approach, a solvent-separated ion pair was located for the catecholate−trimethylsulfonium ion system in aqueous solution at a solute coordinate of about −4.2 Å.[55] Because our primary interest in this work is the description of the TS, we did not extend the exploration of the FES in solution to larger values (in absolute terms) of the solute coordinate. Instead, the enzymatic FES shows a free energy minimum for a contact distance between the cofactor and the substrate, with $q_e^{RS} \approx -1.1$ Å. Interestingly, while the solvent coordinate takes very similar values at the TS in solution and in the enzyme, large differences appear at the reactant regions. In the enzyme, the value of the solvent coordinate in this region is $s_e^{RS} \approx 85$ kcal·mol^{-1}·|e|$^{-1}$, close to the value found for the TS. However, the value of the solvent coordinate in solution is $s_w^{RS} >$ 140 kcal·mol^{-1}·|e|$^{-1}$.

The different behavior of the environment in the catalyzed and noncatalyzed reaction illustrates the concept of enzyme preorganization:[37,70] the electrostatic properties of the active site at the Michaelis complex of COMT are already close to those needed to reach the reaction TS. When going from the reactant complex to the TS in aqueous solution, the solvent

dx.doi.org/10.1021/jp505746x | J. Phys. Chem. B 2015, 119, 873−882

environment must suffer an important reorganization around the two reacting fragments (SAM and catecholate). This reorganization can be quantified from the differences in the solvent coordinate between the TS and the reactant state Δs_i. In aqueous solution this magnitude is $\Delta s_w > 60$ kcal·mol^{-1}·|e|$^{-1}$. Instead, in the enzyme, the reorganization is much smaller $\Delta s_e \approx 18$ kcal·mol^{-1}·|e|$^{-1}$ (note the different scales used for the solvent coordinates in Figure 2 for the enzyme and the solution reaction). This larger reorganization implies a free energy penalty reflected in the larger barrier for the reaction in solution.[55] An interesting feature derived from both FESs is that the environmental reorganization, the changes in the solvent coordinate, mostly precedes the changes taking place along the solute coordinate. According to the average reactive trajectories presented in Figure 2 (depicted as the black continuous line on the FESs), most of the solvent reorganization takes place in the first stages of the reaction. The timing for the evolution of the solute and solvent coordinates is clearer for the reaction in aqueous solution where the change in the solvent coordinate is substantially larger than in the enzyme.

It is interesting to stress the role played by the electrostatic fluctuations of the environment to reduce the free energy barrier. According to the FESs presented in Figure 2, the reaction would present a larger free energy barrier in a frozen environment, especially in aqueous solution. Motions of the solvent/protein facilitate the reaction providing a more adequate electrostatic environment for the progress of the reaction, reducing the free energy barrier. However, the largest fraction of this barrier is still associated with changes along the solute coordinate. Then the chemical process should not be described as environmental motions or vibrations leading to a barrierless methyl transfer scenario. The reaction clearly involves fluctuations of the solute and the solvent degrees of freedom, but the former is in fact the determining step to cross the barrier.

Finally, it is interesting to comment that the qualitative picture obtained from the analysis of the evolution of the chemical process in terms of the solute and the solvent coordinates and the differences observed between the catalyzed reaction and the counterpart process in aqueous solution are quite similar to those obtained for other enzymatic reactions: chalcone isomerase,[46] haloalkane dehalogenase,[35] and dihydrofolate reductase.[71] In terms of the electrostatic potential, the environmental reorganization is drastically reduced in the enzyme with respect to the uncatalyzed reaction, which is translated into a small participation of the solvent in the reaction coordinate. This picture was useful to interpret the experimental kinetic results of isotopically substituted versions of dihydrofolate reductase.[71]

Characterization of the Solvent Coordinate. We have characterized the average frequencies of the solute and solvent coordinates at the TS from the free energy landscape and effective masses derived from the equipartition principle. The average frequencies for the solvent coordinate in aqueous solution and in the enzyme are quite similar, 235 and 225 cm^{-1}, respectively. These values are considerably smaller than the average frequency associated with the solute coordinate, about 750 and 770 cm^{-1} in solution and in the enzyme. Thus, motion along the solvent coordinate is, on average, slower than the motion along the solute coordinate. As described above, in an averaged reactive trajectory solvent motions dominate during the first stages. However, one must take into account that the electrostatic potential is a collective coordinate that involves many fundamental motions or fluctuations of the environment, from fast stretching motions to slow conformational or rotational motions, and then time-separation from solute coordinate motions is not complete. In the case of the enzyme the positively charged Lys144 and the Mg^{2+} ion contribute significantly to the electrostatic potential coordinate, and thus their motions are reflected in the spectrum, while in aqueous solution first solvation shell water molecules contribute more significantly.

Figure 3 shows the autocorrelation function of the solvent coordinate ($\langle\langle \delta s(t)\cdot\delta s(0)\rangle\rangle$) computed from 50 ps TS trajectories

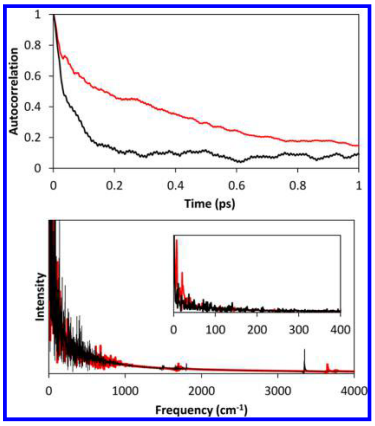


Figure 3. Autocorrelation function (up) and Fourier transform (down) determined from simulations at $q = q^{TS}$ for the system in aqueous solution (red line) and in COMT (black line). The inset focuses on the lower frequencies region of the spectra.

in solution and in COMT where the solute coordinate was kept frozen at $q = q^{TS}$ using a Lagrange-multiplier-based SHAKE algorithm.[72] The time evolution of the autocorrelation function shows a slow decay in the picosecond time scale and also a fine structure reflecting the participation of faster components. Interestingly, a slower decay is observed in solution than in the enzyme. While in principle much slower motions could happen in the enzyme (associated for example with global conformational changes) than in solution, the participation of slow motions in the electrostatic coordinate seems to be more important in solution.

The Fourier spectra of the solvent coordinate in both media are also presented in Figure 3. The electrostatic coordinate is the result of different motions, but most of them are slower than the solute reaction coordinate. In solution we observed a small fast contribution due to the O−H stretching of water molecules hydrogen-bonded to the nucleophile oxygen atom. A larger signal is observed in the enzyme assigned to the N−H stretching of Lys144, a positively charged residue that is hydrogen-bonded to the nucleophilic atom (see Figure 1). Because the reaction proceeds with annihilation of the charge on the substrate, one cannot attribute TS stabilization effects (relative to the reactants) to this hydrogen-bonding interaction. It is important to remember that catalysis is relative to the

dx.doi.org/10.1021/jp505746x | J. Phys. Chem. B 2015, 119, 873−882

Table 1. Average Geometrical Parameters (Distances (D) in Å, Angles (A) in Degrees) and Transmission Coefficients (κ) Obtained for the TS Ensembles Characterized Using the Solute (q) or the Solute and Solvent Coordinates (q,s)

	aqueous solution		COMT	
	$TS_w(q)$	$TS_w(q,s)$	$TS_e(q)$	$TS_e(q,s)$
D(S−C)	2.12 ± 0.04	2.12 ± 0.04	2.17 ± 0.05	2.16 ± 0.05
D(C−O1)	2.03 ± 0.04	2.03 ± 0.05	1.99 ± 0.05	1.98 ± 0.04
D(S−O1)	4.14 ± 0.08	4.11 ± 0.08	4.12 ± 0.09	4.11 ± 0.08
A(S−C−O1)	172 ± 4	167 ± 6	166 ± 5	166 ± 6
D(Mg−O2)			2.23 ± 0.09	2.24 ± 0.10
D(O1−N$^{\varepsilon}$Lys144)			2.74 ± 0.09	2.75 ± 0.09
D(O1−O$_w$)	2.76 ± 0.07	2.73 ± 0.08		
κ	0.49 ± 0.05	0.70 ± 0.05	0.76 ± 0.05	0.72 ± 0.05

process in solution and that rate acceleration can be attained if the TS is less destabilized with respect to the reactants in the enzyme than in solution.[56]

Some low intensity signals appear in both environments in the region between 1400 and 1800 cm^{-1}, corresponding to heavy atoms stretching in the protein and water molecule bending. However, the most intense signals are found below 50 cm^{-1}, which correspond to motions taking place on the picosecond time scale. It must be taken into account that the slower components of the solvent coordinate must precede the motion along the solute coordinate. Only those components of the solvent coordinate presenting higher frequencies could couple to the solute coordinate during the passage through the dividing surface. The enzyme shows more important participation of motions in the 100−500 cm^{-1} range than the aqueous solution, while in aqueous solution some intense signals appear in the 600−700 cm^{-1} range, a frequency similar to that of the solute coordinate. Hydrogen bonds and other electrostatic interactions between the environment and the reacting system appear in these regions of the spectra.

Characterization of TS Ensembles. The TS ensembles defined using only the solute coordinate ($TS(q)$) or on the 2D-surface defined by the solute and the solvent coordinates ($TS(q,s)$) are different, as observed in Figure 2. In the 2D-picture, the explicit consideration of the solvent coordinate produces a rotation of the TS plane. However, it must be taken into account that when the TS is defined exclusively in terms of the solute coordinate, then the solvent is in equilibrium. Thus, $TS(q)$ corresponds to a free energy minimum along the solvent coordinate. Then, except for cases with a very large participation of the solvent coordinate in the barrier crossing and/or with very small curvatures around the saddle point, the average description obtained for the two TS ensembles ($TS(q)$ and $TS(q,s)$) can be quite similar. From a dynamical perspective, the slower components of the solvent reorganization precede changes in the faster solute coordinate, providing an equilibrium-like solvation at the TS. This similarity between both TS ensembles is in fact observed for catecholate methylation in aqueous solution and in COMT. Geometrical parameters given in Table 1 are statistically indistinguishable. The averaged geometrical parameters corresponding to the broken and formed bonds, the donor−acceptor distance, and the attacking angle are similar, with differences below the standard deviations. Slightly larger changes are observed between the TS ensembles in aqueous solution, in particular for the methyl attacking angle (S−C−O1), although the dispersion associated with the structures of the TS ensembles prevent any quantitative conclusion from this observation. We also provide some distances characterizing the interaction of the chemical system to the environment. In the enzyme the distance of the hydroxyl oxygen atom of catechol (O2) to the magnesium ion present in the enzymatic active site and the distance of the nucleophilic oxygen O1 to the ammonium group of Lys144 are equivalent in both TS ensembles. In solution, the average distance of the O1 atom to the closest water molecule (O1−O$_w$) is also equivalent in both ensembles within their standard deviation.

In passing, it is interesting to note that we did not find any evidence for a significant difference in the distances between the methyl donor and acceptor in the enzymatic TS with respect to the solution one. The experimental observation of much more inverse secondary α-deuterium KIEs for the COMT catalyzed reaction than for the uncatalyzed methyl transfer was originally interpreted as the consequence of a tighter transition state in the enzymatic reaction, with significantly smaller donor−acceptor distance.[49] However, QM/MM calculations based on equilibrium simulations along the solute coordinate reproduced the experimental observation without evidence for compression.[58,73] While it could be argued that the sampling resulting from those simulations could be biased to produce similar results irrespective of the environment, the new simulations presented here show that the averaged distances are still the same, within their standard deviations, in both environments. The explicit inclusion of environmental motions in the definition of the TS ensemble does not affect this conclusion, suggesting that the inverse secondary α-deuterium KIES in COMT[32,48,57] can be explained by the preferential equilibrium solvation provided by the environment.[58] The origin of the larger increase of the force constants associated with motions of the α-hydrogens at the enzymatic TS is not due to changes in the interactions between the reacting fragments but in the interactions established with the environments. This finding illustrates the necessity of adequate computational models to assign the physical origin of complex experimental observations. If the environment is neglected, the more inverse KIE must result from the compression of the donor−acceptor distance, but if the environment is included, alternative (and correct) explanations arise.

It is also interesting to compare the TS geometries in solution and in the enzyme in the context of the Hammond postulate.[74] In our simulations a smaller barrier in the enzyme corresponds to a more advanced TS in terms of methyl−donor and methyl−acceptor distances. This behavior can be rationalized considering that the polarity of the solute decreases as the reaction advances, and thus, catalysis is the result of a smaller destabilization of the TS in the enzyme than in solution. This scenario for enzymatic reactions was discussed in ref 56.

dx.doi.org/10.1021/jp505746x | J. Phys. Chem. B 2015, 119, 873−882

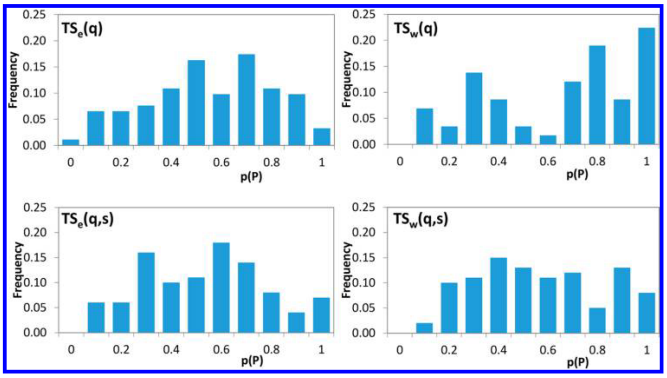


Figure 4. Commitor histograms obtained for the four TS ensembles analyzed from the probabilities that trajectories initiated from a given trajectory commit to products (p(P)). The graphs on the left correspond to the enzymatic reaction and those on the right to the aqueous solution reaction (monodimensional and two-dimensional definitions up and down, respectively).

TS ensembles were also characterized analyzing a set of 5000 trajectories initiated from TS structures selected for each of the four ensembles (50 trajectories started from 100 TS structures for each ensemble, using different velocities taken from a Maxwell−Boltzmann distribution). From these trajectories we computed the transmission coefficient as[11]

$$\kappa = \frac{\sum_{i=1}^{N} v_i Q_i}{\sum_{i=1}^{N} |v_i|} \tag{7}$$

where it has been assumed that all the trajectories have the same probability because they were obtained from an equilibrium distribution. v_i is the initial velocity associated with the reaction coordinate, and Q_i is equal to 1 for reactant to product trajectories, 0 for reactant to reactant or product to product, and −1 for product to reactant trajectories. The values of the transmission coefficient for the four TS ensembles and their standard deviation are provided in Table 1. Standard deviations were obtained according to ref 12 for 100 independent trajectories (the number of initial structures). The values of the transmission coefficient are similar to those obtained in previous studies of the same system using slightly different simulation conditions.[39,41] It is interesting to note that the transmission coefficients calculated from the enzymatic TS ensembles optimized along one (TS$_e$(q)) and two (TS$_e$(q,s)) coordinates are practically indistinguishable, while the transmission coefficient in water increases when passing from the TS$_w$(q) to the TS$_w$(q,s). This effect illustrates the fact that the participation of solvent motions during barrier crossing is significantly more important in aqueous solution than in the enzyme. When solvent motions are incorporated into the definition of the TS ensemble through the electrostatic coordinate, the transmission coefficients are very similar for the catalyzed and the uncatalyzed reaction. This coincidence runs against the proposals about a larger role of dynamical contributions during barrier crossing in enzymatic reactions. Note that the transmission coefficient found for the TS(q,s) ensembles is not unity. First, the existence of a no-recrossing TS hypersurface is not guaranteed even in simple systems because recrossings could emerge not only from a nonoptimal choice of the reaction coordinate but also from nonlinear dynamical effects.[75,76] Second, we have optimized the TS in one and two coordinates, without considering the complete multidimensional configuration space. Other solute and solvent coordinates could have noticeable participation in the accurate determination of the TS ensemble. For example, the methyl inversion was shown to be strongly coupled to the antisymmetric coordinate.[39]

A more stringent test about the quality of the TS ensembles is provided by the analysis of the committor histograms. These histograms, presented in Figure 4, provide the probability that free trajectories initiated from a given TS structure reach the basin of the products before the basin of the reactants. For each of the 100 structures selected for the four TS ensembles given in Table 1, the probability of ending in products for each structure was determined from the set of 50 trajectories. If the TS ensemble is correctly defined, the committor histogram should be a unimodal distribution peaked around 0.5. The broadness observed for the committor distributions presented in Figure 4 can be due to at least three factors: (i) an incomplete optimization of the reaction coordinate (the gradient of the committor function around the TS is usually high); (ii) an insufficient sampling in terms of initial structures and/or trajectories; (iii) dynamical effects.[76]

According to Figure 4, the explicit consideration of the solvent coordinate in the definition of the TS surface in aqueous solution clearly improves the distribution of probabilities to commit to the product basins. Then a more rigorous picture of the TS can be obtained when the solvent coordinate is explicitly considered in aqueous solution. The TS ensemble obtained using exclusively the solute coordinate shows a bimodal distribution that indicates the presence of a small barrier along the solvent coordinate. This barrier can be estimated from the ratio of the transmission coefficients obtained including or not the solvent coordinate. This ratio equals 1.4 and can be translated into a free energy difference of only 0.2 kcal·mol^{-1} at 298 K, below the expected statistical error in the evaluation of activation free energies.

In the enzyme, the consideration of the solvent coordinate in the TS definition moderately improves the committor histo-

dx.doi.org/10.1021/jp505746x | J. Phys. Chem. B 2015, 119, 873−882

gram. This result correlates with the changes observed in the transmission coefficients given in Table 1, which are almost equivalent in the two enzymatic TS ensembles. Participation of environmental motions in the reaction coordinate at the TS seems to be less important than in solution. The larger reorganization needed in solution to reach the TS, reflected in the larger variation observed for the electrostatic coordinate along the reaction (see Figure 2), results in a larger coupling between both coordinates in this medium. It must be emphasized that the solvent coordinate is composed of both fast and slow motions and only those motions presenting time scales similar to those of the solute coordinate can couple during barrier crossing. Slower motions must precede the changes in the solute coordinate, which is reflected in the average reactive trajectory presented in Figure 2.

4. CONCLUSIONS

We have presented a QM/MM computational analysis of catechol O-methylation in aqueous solution and in the active site of COMT. FESs have been obtained as a function of a solute coordinate (the antisymmetric combination of the distances of the methyl group to the donor and the acceptor atoms) and a solvent coordinate (defined as the electrostatic potential created by the environment on the methyl donor and acceptor atoms).

From the analysis of the FESs, shown in Figure 2, and the corresponding TS ensembles we can conclude that the participation of environmental motions in the reaction coordinate at the TS is more important for the reaction in solution than in the enzyme because a larger reorganization of the environment is needed in solution than in the enzyme. The enzymatic active site is largely prepared to accommodate the reaction at the Michaelis complex. During analysis of the changes taking place in the environment during the chemical process, it can be observed that most of the environmental motions participating in the electrostatic reorganization are slower than the solute reaction coordinate. Then a large fraction of the changes in the solvent coordinate takes place during the first stages of the process. However, some of these motions can happen on similar time scales and can couple to the solute coordinate during barrier crossing. This is reflected in a slight rotation of the TS surface obtained in the two-dimensional description with respect to the equilibrium treatment (see Figure 2). However, our results clearly show that the use of the equilibrium approximation results in a very good definition of the TS ensemble in both environments and that a nonequilibrium treatment of the environment does not significantly change the TS ensembles (see the average geometrical parameters given in Table 1).

The effect of an explicit consideration of a solvent coordinate in the definition of the TS ensemble can be appreciated calculating the transmission coefficients and the committor histograms. The inclusion of the solvent coordinate improves the transmission coefficient in aqueous solution by a factor of 1.4. The committor histogram obtained using only the solute coordinate shows that a small free energy barrier is present along the solvent coordinate (about 0.2 kcal·mol^{-1} at 298 K). The changes observed in the transmission coefficients and committor histograms calculated for the enzymatic reaction are significantly less significant than in aqueous solution. This observation agrees with a smaller participation of environmental motions in the catalyzed reaction than in the uncatalyzed one. The equilibrium solvation approach works even better for the enzymatic reaction than for the reaction in solution.

AUTHOR INFORMATION

Corresponding Authors

*J.J.R.-P.: e-mail, pernia@qfa.uji.es.

*I.T.: e-mail, ignacio.tunon@uv.es.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support from FEDER funds and the Ministerio de Economía y Competitividad (Project CTQ2012-36253-C03) and Generalitat Valencia Project GV/2012/053. R.G.-M. and K.Z. acknowledge a fellowship from Ministerio de Educación (FPI and FPU, respectively). The authors acknowledge computational facilities of the Servei d'Informàtica de la Universitat de València for use the "Tirant" supercomputer.

REFERENCES

(1) Jorgensen, W. L.; Ravimohan, C. Monte Carlo Simulation of Differences in Free Energies of Hydration. *J. Chem. Phys.* **1985**, *83*, 3050−3054.

(2) Jorgensen, W. L.; Buckner, J. K.; Huston, S. E.; Rossky, P. J. Hydration and Energetics for *tert*-Butyl Chloride Ion Pairs in Aqueous Solution. *J. Am. Chem. Soc.* **1987**, *109*, 1891−1899.

(3) Gushurst, A. J.; Jorgensen, W. L. Computer-Assisted Mechanistic Evaluation of Organic Reactions. 12. pK_a Predictions for Organic Compounds in Me_2SO. *J. Org. Chem.* **1986**, *51*, 3513−3522.

(4) Jorgensen, W. L. Monte Carlo Simulation of n-Butane in Water. Conformational Evidence for the Hydrophobic Effect. *J. Chem. Phys.* **1982**, *77*, 5757−5765.

(5) Smith, R. H.; Jorgensen, W. L.; Tirado-Rives, J.; Lamb, M. L.; Janssen, P. A. J.; Michejda, C. J.; Kroeger Smith, M. B. Prediction of Binding Affinities for TIBO Inhibitors of HIV-1 Reverse Transcriptase Using Monte Carlo Simulations in a Linear Response Method. *J. Med. Chem.* **1998**, *41*, 5272−5286.

(6) Jorgensen, W. L. Rusting of the Lock and Key Model for Protein−Ligand Binding. *Science* **1991**, *254*, 954−955.

(7) Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. S_N2 Reaction Profiles in the Gas Phase and Aqueous Solution. *J. Am. Chem. Soc.* **1984**, *106*, 3049−3050.

(8) Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. Theoretical Examination of the S_N2 Reaction Involving Chloride Ion and Methyl Chloride in the Gas Phase and Aqueous Solution. *J. Am. Chem. Soc.* **1985**, *107*, 154−163.

(9) Garrett, B. C.; Truhlar, D. G. Criterion of Minimum State Density in the Transition State Theory of Bimolecular Reactions. *J. Chem. Phys.* **1979**, *70*, 1593−1598.

(10) Eyring, H. The Theory of Absolute Reaction Rates. *J. Chem. Soc., Faraday Trans.* **1938**, *34*, 41−48.

(11) Bergsma, J. P.; Gertner, B. J.; Wilson, K. R.; Hynes, J. T. Molecular Dynamics of a Model S_N2 Reaction in Water. *J. Chem. Phys.* **1987**, *86*, 1356−1376.

(12) Gertner, B. J.; Wilson, K. R.; Hynes, J. T. Nonequilibrium Solvation Effects on Reaction Rates for Model S_N2 Reactions in Water. *J. Chem. Phys.* **1989**, *90*, 3537−3558.

(13) Hwang, J. K.; King, G.; Creighton, S.; Warshel, A. Simulation of Free Energy Relationships and Dynamics of S_N2 Reactions in Aqueous Solution. *J. Am. Chem. Soc.* **1988**, *110*, 5297−5311.

(14) Hammes, G. G.; Benkovic, S. J.; Hammes-Schiffer, S. Flexibility, Diversity, and Cooperativity: Pillars of Enzyme Catalysis. *Biochemistry* **2011**, *50*, 10422−10430.

(15) Henzler-Wildman, K. A.; Lei, M.; Thai, V.; Kerns, S. J.; Karplus, M.; Kern, D. A Hierarchy of Timescales in Protein Dynamics is Linked to Enzyme Catalysis. *Nature* **2007**, *450*, 913−916.
(16) Warshel, A. Dynamics of Enzymatic Reactions. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 444−448.
(17) Cameron, C. E.; Benkovic, S. J. Evidence for a Functional Role of the Dynamics of Glycine-121 of *Escherichia coli* Dihydrofolate Reductase Obtained from Kinetic Analysis of a Site-Directed Mutant. *Biochemistry* **1997**, *36*, 15792−15800.
(18) Radkiewicz, J. L.; Brooks, C. L. Protein Dynamics in Enzymatic Catalysis: Exploration of Dihydrofolate Reductase. *J. Am. Chem. Soc.* **1999**, *122*, 225−231.
(19) Antoniou, D.; Schwartz, S. D. Internal Enzyme Motions as a Source of Catalytic Activity: Rate-Promoting Vibrations and Hydrogen Tunneling. *J. Phys. Chem. B* **2001**, *105*, 5553−5558.
(20) Benkovic, S. J.; Hammes-Schiffer, S. A Perspective on Enzyme Catalysis. *Science* **2003**, *301*, 1196−1202.
(21) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. How Enzymes Work: Analysis by Modern Rate Theory and Computer Simulations. *Science* **2004**, *303*, 186−195.
(22) Antoniou, D.; Basner, J.; Núñez, S.; Schwartz, S. D. Computational and Theoretical Methods To Explore the Relation between Enzyme Dynamics and Catalysis. *Chem. Rev.* **2006**, *106*, 3170−3187.
(23) Gao, J.; Ma, S.; Major, D. T.; Nam, K.; Pu, J.; Truhlar, D. G. Mechanisms and Free Energies of Enzymatic Reactions. *Chem. Rev.* **2006**, *106*, 3188−3209.
(24) Olsson, M. H. M.; Parson, W. W.; Warshel, A. Dynamical Contributions to Enzyme Catalysis: Critical Tests of a Popular Hypothesis. *Chem. Rev.* **2006**, *106*, 1737−1756.
(25) Pu, J.; Gao, J.; Truhlar, D. G. Multidimensional Tunneling, Recrossing, and the Transmission Coefficient for Enzymatic Reactions. *Chem. Rev.* **2006**, *106*, 3140−3169.
(26) Kamerlin, S. C. L.; Warshel, A. At the Dawn of the 21st Century: Is Dynamics the Missing Link for Understanding Enzyme Catalysis? *Proteins* **2010**, *78*, 1339−1375.
(27) Nashine, V. C.; Hammes-Schiffer, S.; Benkovic, S. J. Coupled Motions in Enzyme Catalysis. *Curr. Opin. Chem. Biol.* **2010**, *14*, 644−651.
(28) Adamczyk, A. J.; Cao, J.; Kamerlin, S. C. L.; Warshel, A. Catalysis by Dihydrofolate Reductase and Other Enzymes Arises from Electrostatic Preorganization, Not Conformational Motions. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 14115−14120.
(29) Bhabha, G.; Lee, J.; Ekiert, D. C.; Gam, J.; Wilson, I. A.; Dyson, H. J.; Benkovic, S. J.; Wright, P. E. A Dynamic Knockout Reveals That Conformational Fluctuations Influence the Chemical Step of Enzyme Catalysis. *Science* **2011**, *332*, 234−238.
(30) Boekelheide, N.; Salomón-Ferrer, R.; Miller, T. F. Dynamics and Dissipation in Enzyme Catalysis. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 16159−16163.
(31) Ramanathan, A.; Agarwal, P. K. Evolutionarily Conserved Linkage between Enzyme Fold, Flexibility, and Catalysis. *PLoS Biol.* **2011**, *9*, e1001193.
(32) Zhang, J.; Klinman, J. P. Enzymatic Methyl Transfer: Role of an Active Site Residue in Generating Active Site Compaction That Correlates with Catalytic Efficiency. *J. Am. Chem. Soc.* **2011**, *133*, 17134−17137.
(33) Hay, S.; Scrutton, N. S. Good Vibrations in Enzyme-Catalysed Reactions. *Nat. Chem.* **2012**, *4*, 161−168.
(34) Glowacki, D. R.; Harvey, J. N.; Mulholland, A. J. Taking Ockham's Razor to Enzyme Dynamics and Catalysis. *Nat. Chem.* **2012**, *4*, 169−176.
(35) Garcia-Meseguer, R.; Marti, S.; Ruiz-Pernia, J. J.; Moliner, V.; Tuñon, I. Studying the Role of Protein Dynamics in an S_N2 Enzyme Reaction Using Free-Energy Surfaces and Solvent Coordinates. *Nat. Chem.* **2013**, *5*, 566−571.
(36) Doron, D.; Kohen, A.; Nam, K.; Major, D. T. How Accurate Are Transition States from Simulations of Enzymatic Reactions? *J. Chem. Theory Comput.* **2014**, *10*, 1863−1871.
(37) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H. B.; Olsson, M. H. M. Electrostatic Basis for Enzyme Catalysis. *Chem. Rev.* **2006**, *106*, 3210−3235.
(38) Karplus, M.; McCammon, J. A. Dynamics of Proteins: Elements and Function. *Annu. Rev. Biochem.* **1983**, *52*, 263−300.
(39) Roca, M.; Moliner, V.; Tuñón, I.; Hynes, J. T. Coupling between Protein and Reaction Dynamics in Enzymatic Processes: Application of Grote-Hynes Theory to Catechol *O*-Methyltransferase. *J. Am. Chem. Soc.* **2006**, *128*, 6186−6193.
(40) Garcia-Viloca, M.; Truhlar, D. G.; Gao, J. Reaction-Path Energetics and Kinetics of the Hydride Transfer Reaction Catalyzed by Dihydrofolate Reductase. *Biochemistry* **2003**, *42*, 13558−13575.
(41) Roca, M.; Andrés, J.; Moliner, V.; Tuñón, I.; Bertrán, J. On the Nature of the Transition State in Catechol *O*-Methyltransferase. A Complementary Study Based on Molecular Dynamics and Potential Energy Surface Explorations. *J. Am. Chem. Soc.* **2005**, *127*, 10648−10655.
(42) Whittier, S. K.; Hengge, A. C.; Loria, J. P. Conformational Motions Regulate Phosphoryl Transfer in Related Protein Tyrosine Phosphatases. *Science* **2013**, *341*, 899−903.
(43) Kosugi, T.; Hayashi, S. Crucial Role of Protein Flexibility in Formation of a Stable Reaction Transition State in an α-Amylase Catalysis. *J. Am. Chem. Soc.* **2012**, *134*, 7045−7055.
(44) Warshel, A.; Weiss, R. M. An Empirical Valence Bond Approach for Comparing Reactions in Solutions and in Enzymes. *J. Am. Chem. Soc.* **1980**, *102*, 6218−6226.
(45) Ruiz-Pernia, J. J.; Tuñón, I.; Moliner, V.; Hynes, J. T.; Roca, M. Dynamic Effects on Reaction Rates in a Michael Addition Catalyzed by Chalcone Isomerase. Beyond the Frozen Environment Approach. *J. Am. Chem. Soc.* **2008**, *130*, 7477−7488.
(46) Ruiz-Pernía, J. J.; Martí, S.; Moliner, V.; Tuñón, I. A Novel Strategy To Study Electrostatic Effects in Chemical Reactions: Differences between the Role of Solvent and the Active Site of Chalcone Isomerase in a Michael Addition. *J. Chem. Theory Comput.* **2012**, *8*, 1532−1535.
(47) Takusagawa, F.; Fujioka, M.; Spies, A.; Schowen, R. L. In *Comprehensive Biological Catalyzis*; Sinnott, M., Ed.; Academic Press: San Diego, CA, 1998; Vol. *1*, pp 1−30.
(48) Hegazi, M. F.; Borchardt, R. T.; Schowen, R. L. alpha.-Deuterium and Carbon-13 Isotope Effects for Methyl Transfer Catalyzed by Catechol *O*-Methyltransferase. S_N2-like Transition State. *J. Am. Chem. Soc.* **1979**, *101*, 4359−4365.
(49) Rodgers, J.; Femec, D. A.; Schowen, R. L. Isotopic Mapping of Transition-State Structural Features Associated with Enzymic Catalysis of Methyl Transfer. *J. Am. Chem. Soc.* **1982**, *104*, 3263−3268.
(50) Woodard, R. W.; Tsai, M. D.; Floss, H. G.; Crooks, P. A.; Coward, J. K. Stereochemical Course of the Transmethylation Catalyzed by Catechol *O*-Methyltransferase. *J. Biol. Chem.* **1980**, *255*, 9124−9127.
(51) Knipe, J. O.; Vasquez, P. J.; Coward, J. K. Methylase Models: Studies on General-Base vs. Nucleophilic Catalysis in the Intramolecular Alkylation of Phenols. *J. Am. Chem. Soc.* **1982**, *104*, 3202−3209.
(52) Zheng, Y.; Bruice, T. C. A Theoretical Examination of the Factors Controlling the Catalytic Efficiency of a Transmethylation Enzyme: Catechol *O*-Methyltransferase. *J. Am. Chem. Soc.* **1997**, *119*, 8137−8145.
(53) Kahn, K.; Bruice, T. C. Transition-State and Ground-State Structures and Their Interaction with the Active-Site Residues in Catechol *O*-Methyltransferase. *J. Am. Chem. Soc.* **1999**, *122*, 46−51.
(54) Lau, E. Y.; Bruice, T. C. Comparison of the Dynamics for Ground-State and Transition-State Structures in the Active Site of Catechol *O*-Methyltransferase. *J. Am. Chem. Soc.* **2000**, *122*, 7165−7171.
(55) Roca, M.; Martí, S.; Andrés, J.; Moliner, V.; Tuñón, I.; Bertrán, J.; Williams, I. H. Theoretical Modeling of Enzyme Catalytic Power: Analysis of "Cratic" and Electrostatic Factors in Catechol *O*-Methyltransferase. *J. Am. Chem. Soc.* **2003**, *125*, 7726−7737.

(56) Tuñón, I.; Hynes, J. T. A Simple Model for Barrier Frequencies for Enzymatic Reactions. *ChemPhysChem* **2011**, *12*, 184−190.
(57) Gray, C. H.; Coward, J. K.; Schowen, K. B.; Schowen, R. L. alpha.-Deuterium and Carbon-13 Isotope Effects for a Simple, Intermolecular Sulfur-to-Oxygen Methyl-Transfer Reaction. Transition-State Structures and Isotope Effects in Transmethylation and Transalkylation. *J. Am. Chem. Soc.* **1979**, *101*, 4351−4358.
(58) Ruggiero, G. D.; Williams, I. H.; Roca, M.; Moliner, V.; Tuñón, I. QM/MM Determination of Kinetic Isotope Effects for COMT-Catalyzed Methyl Transfer Does Not Support Compression Hypothesis. *J. Am. Chem. Soc.* **2004**, *126*, 8634−8635.
(59) Roca, M.; Moliner, V.; Ruiz-Pernía, J. J.; Silla, E.; Tuñón, I. Activation Free Energy of Catechol *O*-Methyltransferase. Corrections to the Potential of Mean Force. *J. Phys. Chem. B* **2005**, *110*, 503−509.
(60) Vidgren, J.; Svensson, L. A.; Liljas, A. Crystal Structure of Catechol *O*-Methyltransferase. *Nature* **1994**, *368*, 354−358.
(61) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and Use of Quantum Mechanical Molecular Models. 76. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(62) Ruiz-Pernía, J. J.; Silla, E.; Tuñón, I.; Martí, S. Hybrid Quantum Mechanics/Molecular Mechanics Simulations with Two-Dimensional Interpolated Corrections: Application to Enzymatic Processes. *J. Phys. Chem. B* **2006**, *110*, 17663−17670.
(63) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(64) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(65) Neria, E.; Fischer, S.; Karplus, M. Simulation of Activation Free Energies in Molecular Systems. *J. Chem. Phys.* **1996**, *105*, 1902−1921.
(66) Nam, K.; Gao, J.; York, D. M. An Efficient Linear-Scaling Ewald Method for Long-Range Electrostatic Interactions in Combined QM/MM Calculations. *J. Chem. Theory Comput.* **2004**, *1*, 2−13.
(67) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.
(68) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(69) Kästner, J. Umbrella Integration in Two or More Reaction Coordinates. *J. Chem. Phys.* **2009**, *131*, 034109.
(70) Warshel, A. Electrostatic Origin of the Catalytic Power of Enzymes and the Role of Preorganized Active Sites. *J. Biol. Chem.* **1998**, *273*, 27035−27038.
(71) Ruiz-Pernia, J. J.; Luk, L. Y. P.; Garcia-Meseguer, R.; Martí, S.; Loveridge, E. J.; Tuñón, I.; Moliner, V.; Allemann, R. K. Increased Dynamic Effects in a Catalytically Compromised Variant of *Escherichia coli* Dihydrofolate Reductase. *J. Am. Chem. Soc.* **2013**, *135*, 18689−18696.
(72) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of *n*-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(73) Kanaan, N.; Ruiz-Pernia, J. J.; Williams, I. H. QM/MM Simulations for Methyl Transfer in Solution and Catalysed by COMT: Ensemble-Averaging of Kinetic Isotope Effects. *Chem. Commun.* **2008**, 6114−6116.
(74) Hammond, G. S. A Correlation of Reaction Rates. *J. Am. Chem. Soc.* **1955**, *77*, 334−338.
(75) Mullen, R. G.; Shea, J. E.; Peters, B. Transmission Coefficients, Committors, and Solvent Coordinates in Ion-Pair Dissociation. *J. Chem. Theory Comput.* **2014**, *10*, 659−667.
(76) Mullen, R. G.; Shea, J. E.; Peters, B. Communication: An Existence Test for Dividing Surfaces without Recrossing. *J. Chem. Phys.* **2014**, *140*, 041104.

Resumen

Introducción

Las enzimas participan en muchos de los procesos biológicos que se dan en organismos vivos.[1-4] Sin ellas la mayoría de las reacciones necesarias para la vida tardarían demasiado tiempo o las condiciones necesarias para que ocurrieran serían demasiado extremas. La capacidad de las enzimas para catalizar reacciones en condiciones relativamente "suaves" las convierte en un potencial objetivo de estudio para el desarrollo de catalizadores industriales, ya que muchas reacciones llevadas a cabo en la industria pueden ser catalizadas por enzimas a temperatura y presión ambiente y con mejor rendimiento.[5,6] El desarrollo de muchos medicamentos se basa en la síntesis de inhibidores de las enzimas que participan en esos procesos necesarios para la vida de patógenos o células cancerosas.[7] Es por eso que, el conocimiento de los mecanismos que envuelven la catálisis enzimática permite el desarrollo de mejores catalizadores y medicamentos a un menor coste.

Aunque las técnicas experimentales nos han ofrecido una gran cantidad de información sobre los procesos enzimáticos, existen muchos aspectos que siguen siendo objeto de debate y que carecen de una respuesta definitiva. Los modelos computacionales han demostrado ser un poderoso aliado a la hora de describir procesos enzimáticos y dar respuesta a muchas de esas preguntas. Mediante el cálculo de los perfiles

de energía libre, los efectos cinéticos isotópicos, las interacciones específicas enzima-sustrato y otras propiedades importantes, la química teórica ha explicado muchos aspectos del comportamiento enzimático que no estaban totalmente resueltos. Los métodos QM/MM permiten incorporar de una forma natural el efecto de entorno sobre la reactividad química tratando adecuadamente todos los átomos presentes en el sistema y con un coste computacional suficientemente bajo como para poder calcular múltiples configuraciones del sistema.[8,9] Éstos métodos dividen el sistema en dos partes: la más grande que incluye la mayor parte de los átomos del entorno y es calculada mediante potenciales clásicos de mecánica molecular (MM), y la parte más pequeña en la que se incluyen aquellos átomos que participan de forma directa en la reacción química y sufren una reorganización electrónica; esta parte es calculada mediante la mecánica quántica (QM). El hamiltoniano total del sistema se obtiene entonces como la suma de los hamiltonianos de cada una de las partes (H_{QM} y H_{MM}) más el hamiltoniano de interacción entre ambas ($H_{QM/MM}$).

$$H_{Total} = H_{QM} + H_{MM} + H_{QM/MM} \qquad (1)$$

Esta metodología nos permite llevar a cabo las simulaciones necesarias para conocer la evolución del sistema durante el proceso catalítico. Pero para poder describir una reacción es necesario definir una coordenada de reacción que indique la situación del sistema en cada momento. Tradicionalmente estas coordenadas han sido definidas como coordenadas geométricas que afectan a unos pocos átomos involucrados directamente en la reacción. Aun así, para poder introducir los efectos que los grados de libertad del entorno (el sistema proteico) tienen sobre

el transcurso de la reacción es más adecuado usar una variable colectiva que puede ser complementaria a la variable geométrica o directamente la única coordenada de reacción. Una variable de este tipo implica a todo el sistema o, al menos, a una gran parte de él y permite seguir la evolución de la reacción química y todo su entorno al mismo tiempo.

El problema de la dinámica de proteínas

Es bien sabido que, para poder funcionar, las enzimas deben ser lo suficientemente flexibles para poder pasar a través de las diferentes conformaciones relevantes en cada paso del proceso catalítico. [10,11] Incluso dentro de un sitio activo diseñado para dar cabida a la distribución de cargas del Estado de Transición (TS),[12] se requieren ciertos movimientos en las coordenadas de la proteína para llegar al mismo desde el complejo reactivo. [11,13-15]

El impacto de la dinámica de proteínas en la constante de velocidad de la etapa química sigue siendo objeto de debate en la literatura científica[14-27]. La cuestión que se plantea es si el efecto de los movimientos de la proteína sobre la constante de velocidad puede ser descrito con un marco teórico basado en la Teoría del estado de Transición (TST).[28,29] Esta teoría proporciona las herramientas para el cálculo de la velocidad de la reacción, sin tener en cuenta explícitamente los efectos dinámicos. Sin embargo, es fácil imaginar que éstos podrían desempeñar un papel importante. De hecho, los efectos dinámicos pueden encontrarse dentro de la propia región de TS. Por ejemplo, un entorno con fuerzas dinámicas viscosas puede limitar ciertos movimientos necesarios para completar la reacción. Transferencias y redistribuciones de carga pueden estar restringidas por el lento movimiento de moléculas polares en el entorno.[24,25]

Un criterio muy común para identificar la presencia de efectos dinámicos es la desviación de la constante de velocidad de su valor obtenido en la TST, [30] medida por el coeficiente de transmisión (κ) y definida por el ratio

k/k_{TST} , siendo k la constante de velocidad de la reacción y k_{TST} la constante de velocidad calculada usando la TST. En este modelo, cuanto más pequeño sea κ, mayor serán los efectos dinámicos sobre la constante de velocidad. No obstante, κ depende también de la definición de la coordenada de reacción (RC) en las proximidades del TS.[31] Por eso, el coeficiente de transmisión podría ser uno (o muy cerca de uno) para una definición de la coordenada de reacción, pero muy pequeño para otra, con una diferencia correspondiente en la evaluación de la importancia de los efectos dinámicos. Otro posible problema surge cuando κ es menor que uno, pero no existe dinámica de entorno, de hecho, en ocasiones, la ausencia de dinámica es responsable de la reducción en κ.[32] Habitualmente, en cualquier sistema condensado se asume que el entorno se encuentra en equilibrio con respecto a la distribución de carga instantánea del soluto reactivo. Esta aproximación de "equilibrio de solvatación" y cualquier desviación de ella produce una reducción del valor de κ.

Para un mejor entendimiento de los efectos dinámicos, necesitamos controlar la evolución del entorno y su comportamiento a lo largo de la transformación química. En estos términos proponemos dos diferentes soluciones:

- Una opción es proyectar la superficie multidimensional de energía libre de la reacción enzimática en un modelo 2D obtenido como función de dos coordenadas, una de soluto y otra de entorno. Una superficie de energía libre como ésta nos permitirá

estimar el acoplamiento entre el soluto y los movimientos del solvente a lo largo de la reacción.

- La segunda solución es definir la RC en términos de las coordenadas de entorno, asumiendo que las partículas ligeras involucradas en la reacción se adaptan al entorno, permitiéndonos la inclusión de los efectos cuánticos en el cálculo de la constante de velocidad.

Coordenada Electrostática

El objetivo de esta tesis ha sido el desarrollo e implementación de dos coordenadas colectivas. Una de ellas es, la coordenada electrostática (Ep), que toma el potencial electrostático, creado por la parte MM en un sistema QM/MM, sobre un átomo directamente involucrado en el proceso de reacción, normalmente uno que sufre un cambio significativo de distribución de carga a lo largo de la reacción química.

$$Ep = V_A = \sum_{j=1}^{N} \frac{q_j}{|\boldsymbol{r_j} - \boldsymbol{r_A}|} \tag{2}$$

Donde A hace referencia al átomo seleccionado del sistema reactivo, q_j es la carga del átomo j, $\boldsymbol{r_j}$ y $\boldsymbol{r_A}$ son las coordenadas de los respectivos átomos.

La coordenada Ep puede ser definida como combinación del potencial electrostático sobre diferentes átomos ya que, cuando un átomo sufre un cambio significativo en su distribución de carga, suele haber otro que sufre la redistribución opuesta. Por eso, una combinación antisimétrica sobre dos átomos (A, B) o grupo de átomos, puede representar mejor el efecto del entorno sobre la reacción.

$$Ep = V_A - V_B = \sum_{j=1}^{N} \frac{q_j}{|\boldsymbol{r_j} - \boldsymbol{r_A}|} - \sum_{j=1}^{N} \frac{q_j}{|\boldsymbol{r_j} - \boldsymbol{r_B}|} \tag{3}$$

Esta metodología nos permitió estudiar los efectos del entorno y su importancia durante el paso químico de la reacción en tres sistemas

diferentes, la Haloalkano Dehalogenasa, la Catecol O-metiltransferasa y la Dihidrofolato Reductasa.

Haloalkano Dehalogenasa

La Haloalkano Dehalogenasa producida por la batería *Xanthobacter autotrophicus* (DhlA), cataliza la hidrólisis del enlace cloro carbono de una gran variedad de compuestos halogenados (ver Figura 1). [33,34]

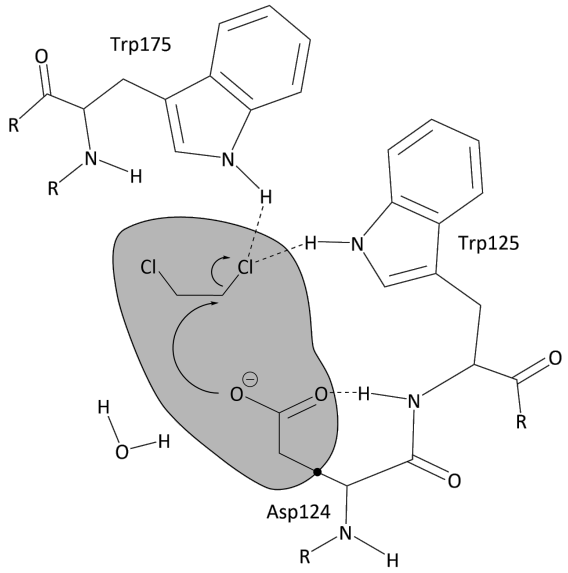


Figura 1: *Representación esquemática de la reacción catalizada por la DhlA. La parte sombreada se corresponde con la parte QM y el punto negro representa el átomo frontera.*

Las superficies de energía libre para la reacción catalizada y la equivalente en agua se obtuvieron como función de dos coordenadas:

- La combinación antisimétrica de las distancias del cloro saliente y del oxígeno entrante al átomo de carbono:

$$q = d(ClC) - d(OC) \qquad (4)$$

- La combinación antisimétrica del potencial electrostático creado por el entorno sobre el cloro saliente y el oxígeno entrante:

$$Ep = V_{CL}(\boldsymbol{r^N}) - V_O(\boldsymbol{r^N}) \quad (5)$$

Como se puede apreciar en la Figura 2 tanto en el sistema en agua como en la enzima la reacción cruza la superficie divisoria a un valor de la coordenada *Ep* similar, sin embargo el complejo reactivo se encuentra a valores mucho más negativos de esta coordenada para el sistema en agua, por lo que el entorno debe sufrir una mayor reordenación para llegar al TS.

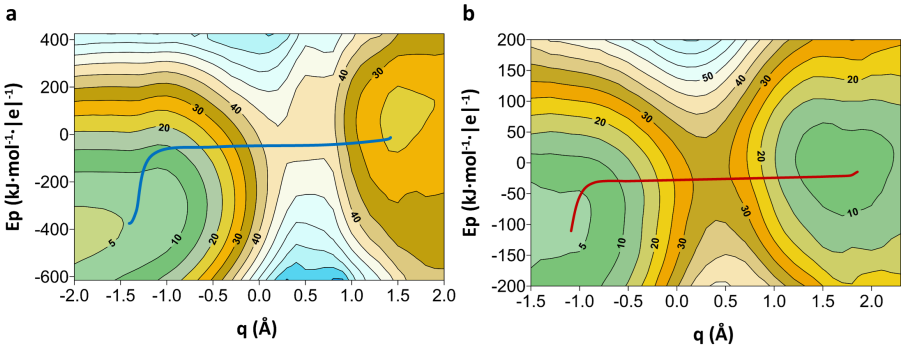


Figura 2: *Superficie de energía libre de la reacción Sn2 de dehalogenación de dicloroetano para la reacción no catalizada (*a*) y en el centro activo Dhla (*b*). Las líneas isoenergéticas están dibujadas cada* $3\ kcal \cdot moL^{-1}$ *Y las líneas azul y roja representan el camino de mínima energía libre en la superficie.*

Podemos separar la barrera energética en su componente a lo largo de la coordenada de entorno (*Ep*) y su componente a lo largo de la coordenada de soluto (q):

$$\Delta G^{\ddagger} \approx \Delta G_s(s_{RS} \rightarrow s^{\ddagger}, q) + \Delta G_q(s, q_{RS} \rightarrow q^{\ddagger}) \qquad (6)$$

donde $\Delta G_s(s_{RS} \rightarrow s^{\ddagger}, q)$ representa el trabajo necesario a lo largo de la coordenada del solvente y $\Delta G_q(s, q_{RS} \rightarrow q^{\ddagger})$ el trabajo en la coordenada de soluto. En el caso de la componente $\Delta G_q(s, q_{RS} \rightarrow q^{\ddagger})$ tanto para el agua como para la enzima tenemos un valor muy similar (27 $kcal \cdot mol^{-1}$ para el agua y 25 $kcal \cdot mol^{-1}$ para la enzima) mientras que en la componente $\Delta G_s(s_{RS} \rightarrow s^{\ddagger}, q)$ encontramos una diferencia mucho más significativa, (11 $kcal \cdot mol^{-1}$ en agua frente a 3 $kcal \cdot mol^{-1}$ de la enzima). Es decir, la reducción en la barrera de energía libre puede atribuirse fundamentalmente al trabajo que ha de hacer el entorno acuoso para adaptarse al TS, mientras que el entorno enzimático se encuentra ya correctamente preorganizado en el entorno de reactivos.

El cálculo de las frecuencias asociadas a los movimientos a lo largo de la coordenada de solvente nos muestra que la estructura de la proteína está mucho más restringida que la estructura del agua, ya que las constate de fuerza obtenida en el sistema enzimático es unas 4.2 veces más grande que la obtenida en agua. Sin embargo, la masa efectiva es también mayor en la enzima, lo que compensa el efecto de la constante de fuerza dando como resultado unas frecuencias similares para ambos sistemas. Esto significa que el tipo de movimientos implicados en el paso de reactivos a TS es similar en ambos entornos y no parece que pueda haber diferencias significativas en la dinámica del mismo entre el agua y la enzima.

Catecol O-Metiltransferasa

La enzima catecol O-metiltransferasa (COMT) cataliza la transferencia de metilo desde S-adenosilmetionina (SAM) al oxígeno nucleofílico del catecol (ver Figura 3).[35-37]

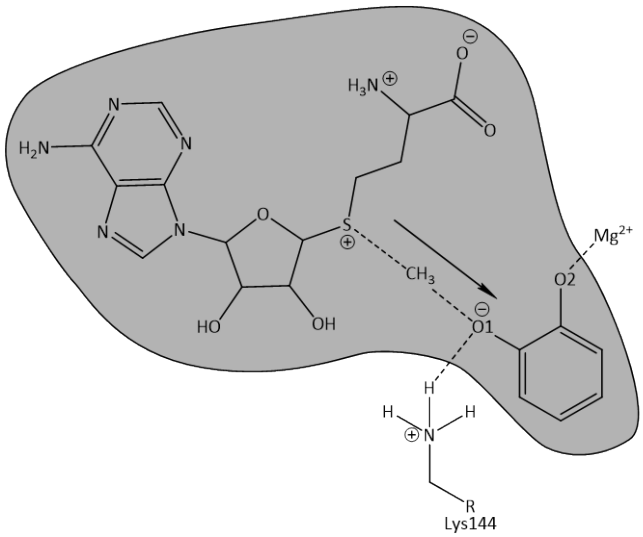


Figura 3: *Representación esquemática de la reacción de metilación catalizada por la COMT. La parte sombreada corresponde a la parte QM.*

Al igual que en el caso de la DhlA las superficies de energía libre para la reacción catalizada y su equivalente en agua se obtuvieron en función de una coordenada de soluto y una de solvente:

- La combinación antisimétrica de las distancias del átomo de azufre del SAM y el oxígeno del catecol al átomo de carbono del metilo transferido:

$$q = d(SC) - d(O1C) \quad (7)$$

- La combinación antisimétrica del potencial electrostático creado por el entorno sobre el azufre del SAM y el oxígeno del catecol:

$$Ep = V_{O1}(\boldsymbol{r^N}) - V_S(\boldsymbol{r^N}) \quad (8)$$

La Figura 4 muestra una variación mayor a lo largo de la coordenada Ep en comparación con las superficies obtenidas en el sistema anterior, tanto para la enzima como para el agua. Esto es debido a que durante la reacción tiene lugar una neutralización de las cargas localizadas sobre el azufre y sobre el oxígeno, lo que da lugar a un cambio mucho mayor en el potencial electrostático sentido por estos átomos. Este cambio sigue siendo mucho mayor en el sistema acuoso que en la enzima. Si definimos la superficie divisoria sobre la superficie de energía libre en función de las dos coordenadas, suponiendo un solvente no equilibrado, vemos una diferencia significativa con respecto a la definición usando solo la coordenada de soluto, suponiendo que el solvente está en equilibrio (ver Figura 4), siendo esta diferencia mucho más grande en agua. Sin embargo, los mínimos de energía en ambos conjuntos coinciden en el mismo punto por lo que no debería haber una diferencia significativa en el valor de la barrera de energía libre.

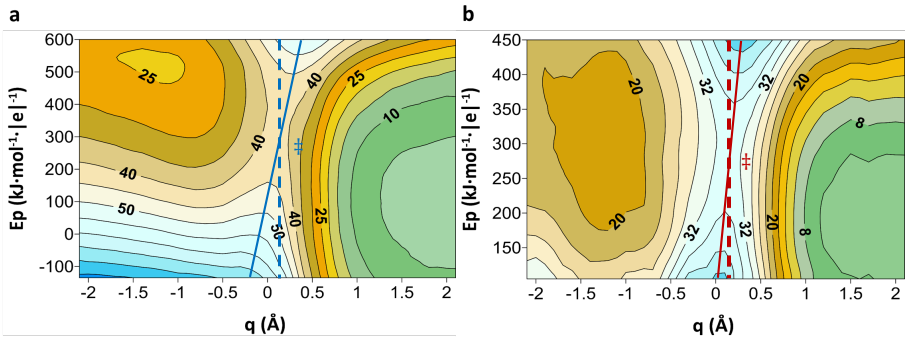


Figura 4: *Superficies de energía libre de la trasferencia de metilo desde S-adenosilmetionina (SAM) al oxígeno nucleofílico del catecol, en agua (*a*) y en la enzima (*b*). Las líneas isoenergéticas están dibujadas cada* 5 $kcal \cdot mol^{-1}$*. Las líneas continuas azul y rojas representan el conjunto de TS obtenido usando la definición de no-equilibrio de solvente. Mientras que las líneas discontinuas representan el conjunto de TS obtenido usando la definición de equilibrio de solvente.*

Aun así, la definición del conjunto de TS usando la aproximación de equilibrio puede dar lugar al encuentro de barreras de energía, en trayectorias que ya han cruzado la superficie divisoria, produciendo recruzamientos y por lo tanto reduciendo el valor de κ.

Para calcular el efecto de la inclusión de las coordenadas de solvente en la definición del TS y en el valor de κ realizamos diversas trayectorias en las diferentes definiciones del mismo: usando las coordenadas de soluto y de solvente para el agua ($TS_w(q, Ep)$) y para la enzima ($TS_e(q, Ep)$) y usando solo la coordenada de soluto para el agua ($TS_w(q)$) y para la enzima ($TS_e(q)$). A continuación, a partir de las trayectorias se extrajeron 100 estructuras de TS para cada una de las definiciones. Por cada estructura se llevaron a cabo 50 trayectorias libres con el objetivo de calcular κ y el histograma del commitor; definido como la probabilidad de

una estructura de TS de acabar en productos; lo que nos permitirá estimar la importancia de incluir las coordenadas de entorno en la definición del TS.

Al calcular κ y el histograma del commitor (ver Figura 5) vemos una significativa mejora al incluir la coordenada de entorno en el sistema en agua, mientras que en el sistema enzimático no existen diferencias significativas. Estos resultados muestran que los efectos dinámicos son más importantes en la reacción en agua, mientras que la enzima reduce significativamente la relevancia de estos efectos.

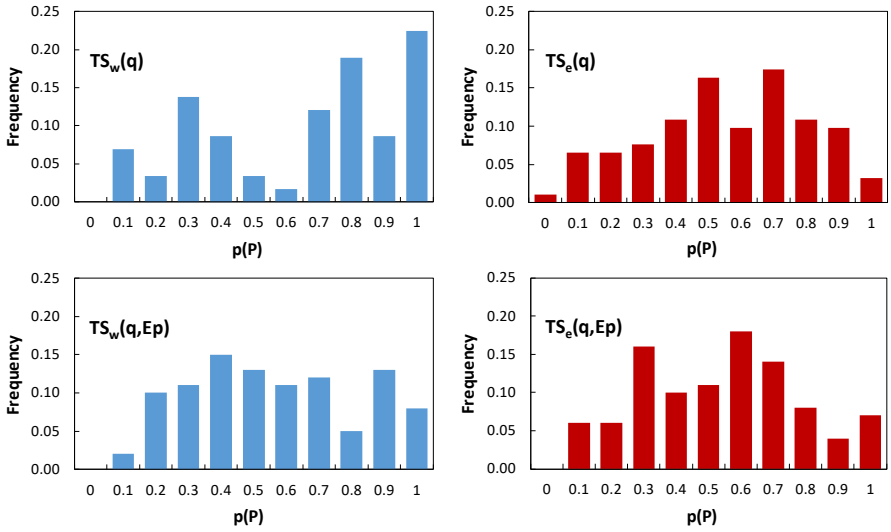


Figura 5: *Histograma de commitor obtenidos de las cuatro diferentes definiciones de TS: unidimensional (arriba) y bidimensional (abajo). Analizados a partir de la probabilidad de una trayectoria dada de acabar en productos (p(P)). Las gráficas azules corresponden a la reacción en agua mientras que las rojas corresponden a la reacción en la enzima.*

Dihidrofolato Reductasa

La enzima dihidrofolato reductasa (DHFR) cataliza la reducción del 7,8-dihidrofolato a 5,6,7,8-tetrahidrofolato mediante la transferencia de hidruro desde el cofactor NADPH (ver Figura 6). [38]

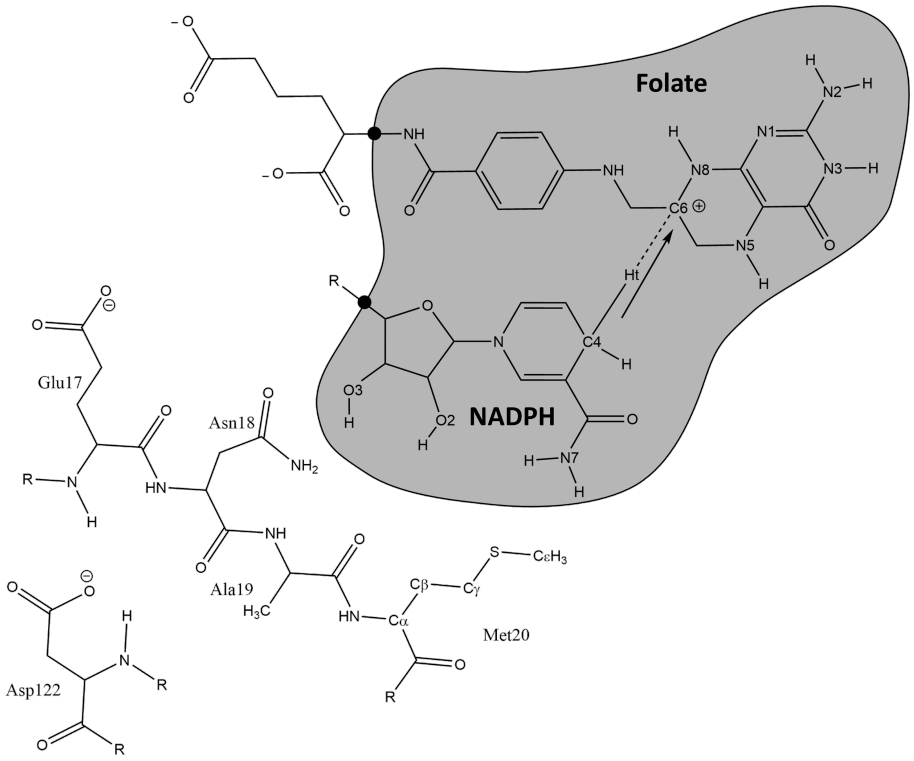


Figura 6: *Representación esquemática de la reacción de transferencia de hidruro catalizada por la DHFR. La parte sombreada se corresponde con la parte QM y los puntos negros representan los átomos frontera. Las etiquetas de cada átomo se usan como referencia en el texto.*

En este caso, se llevaron a cabo tanto cálculos teóricos como experimentos en colaboración con el grupo de Rudolf K. Allemann.[26] Los cálculos estuvieron centrados en el estudio de los efectos dinámicos que

ocurren en el bucle Met20 (ver Figura 7), una sección de la enzima muy importante en el proceso enzimático.[38-41] Este bucle ha sido el objeto de varia mutaciones dirigidas con el objetivo de comprender la importancia de su flexibilidad.[20,21,26,39,42-45]

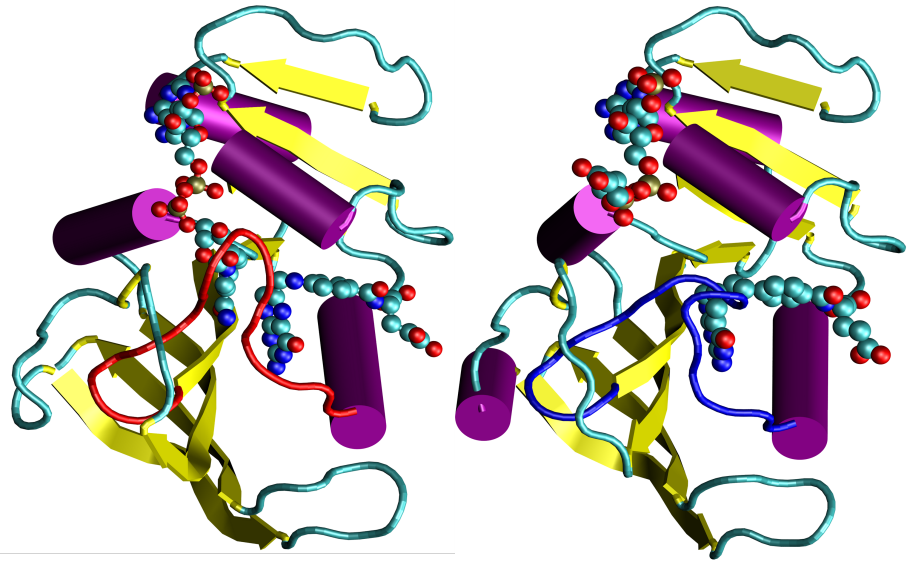


Figura 7: *Representación tridimensional de la DHFR. El bucle Met20 se encuentra pintado de color rojo en su conformación cerrada (izquierda) y en color azul en la conformación ocluida (derecha).*

En concreto nos centraremos en el caso de la mutante EcDHFR-N23PP/S148A, ya que ha levantado cierta controversia sobre el papel de los movimientos enzimáticos en las reacciones de transferencia de hidruro. [20,21,46] En la mutante el bucle Met20 no puede adoptar la conformación ocluida debido a la ausencia de un enlace de hidrógeno crucial para ello. Además, movimientos en escala de los mili- a microsegundos observados en el bucle Met20 desaparecen.[21] Esta mutante muestra una constante de velocidad menor para la transferencia

de hidruro, por lo que se asumió que los movimientos perdidos en el bucle de la mutante jugaban un papel fundamental en la reacción de transferencia en la EcDHFR original. [21] Sin embargo, estudios teóricos basados en la aproximación empírica de enlace de valencia (EVB) sugieren que la reducción de la contante de velocidad es en realidad debida a los efectos electrostáticos y por consiguiente a la energía libre de reorganización en el sitio activo.[20]

Las pruebas experimentales y teóricas realizadas consistieron en comparar las versiones ligera (la enzima normal) y pesada (la misma enzima con los átomos de nitrógeno, carbono y los de hidrógeno no intercambiables sustituidos por sus isotopos pesados ^{15}N, ^{13}C y ^{2}H) de la mutante EcDHFR-N23PP/S148A y comparar los resultados con los obtenidos en trabajos similares previos sobre la EcDHFR original.[47]

Los resultados experimentales muestran en la mutante un KIE enzimático (el cociente entre la constante de velocidad de las enzimas ligera y pesada, k_{LE}/k_{HE}) mayor que los obtenidos para la enzima original.[47] Estos resultados son reproducidos adecuadamente en las simulaciones teóricas y sugieren una mayor influencia de los efectos dinámicos en la mutante EcDHFR-N23PP/S148A que en la enzima original.

Para profundizar en la importancia de los movimientos de la proteína procedimos a calcular las superficies de energía libre; para la enzima EcDHFR, su mutante EcDHFR-N23PP/S148A y para la reacción equivalente en agua(ver Figura 8); usando coordenadas de soluto y solvente como en los sistemas previos:

- La combinación antisimétrica de las distancias del carbono dador del NADH($C4$) y el carbono aceptor del folato (C6) al hidruro transferido:

$$q = d(C4H) - d(C6H) \quad (9)$$

- La combinación antisimétrica del potencial electrostático creado por el entorno sobre el carbono dador ($C4$) y el carbono aceptor (C6).

$$Ep = V_{C6}(\boldsymbol{r^N}) - V_{C4}(\boldsymbol{r^N}) \quad (10)$$

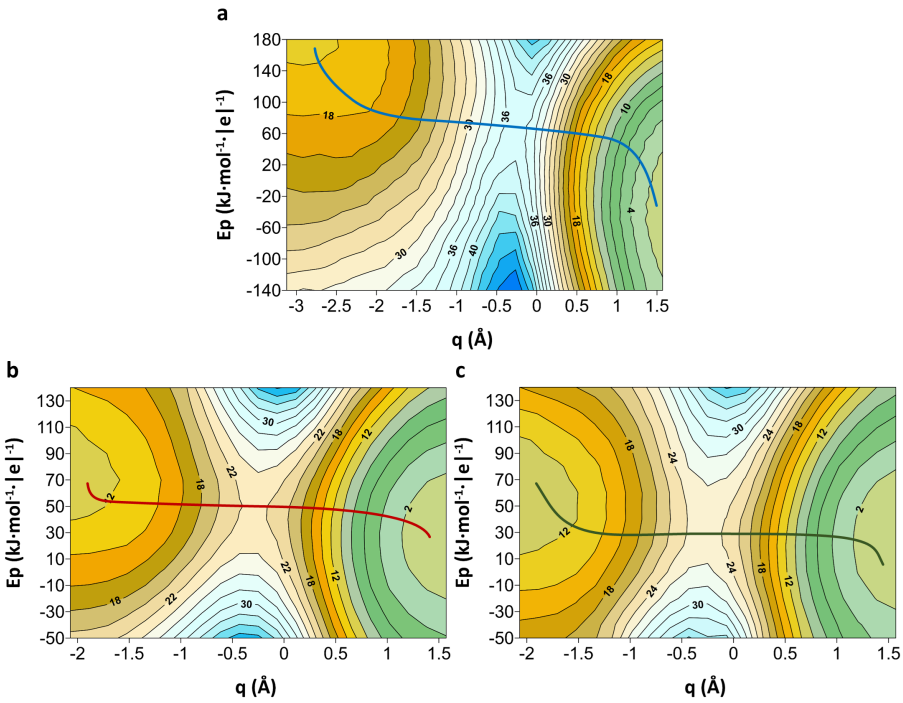


Figura 8: *Superficies de energía libre de la reacción de reducción del 7,8-dihydrofolato correspondientes a la solución acuosa (a), la EcDHFR (b) y la mutante EcDHFR-N23PP/S148A (c). Las líneas isoenergéticas están dibujadas cada 2 kcal · mol*$^{-1}$*. Las líneas continuas representan el camino de mínima energía libre obtenido del gradiente de las superficies.*

Siguiendo los caminos de mínima energía libre se puede ver que el sistema acuoso necesita un mayor reordenamiento del entorno cuando lo comparas con los sistemas enzimáticos. Sin embargo, al calcular las frecuencias de la coordenada del solvente en los tres sistemas éstas no cambian significativamente. Esto no se corresponde con una mutación que afecte directamente a la dinámica de la proteína que afectara al paso químico del proceso enzimático. Por eso, aunque no se excluyen los efectos de la mutación sobre los movimientos de la proteína en la escala

de los milisegundos,[21] estos cambios no afectan directamente a la barrera energética durante la transferencia de hidruro. Sin embargo, en la mutante existen una serie de interacciones entre el cofactor y los residuos asociados al bucle Met20 que han sido alteradas respecto a la enzima original perturbando el proceso de transferencia de hidruro y aumentando su barrera de energía libre.

Coordenada de Diferencia Energética

La segunda coordenada colectiva, que llamaremos Coordenada de Diferencia Energética ($\Delta\varepsilon$)**,** está diseñada para el cálculo de reacciones enzimáticas en las que tiene lugar la transferencia de una partícula ligera como un protón o un hidruro. La definición de esta coordenada se basa en la diferencia de energía entre dos estados de referencia de "Reactivos" (R) y "Productos" (P) definidos en base a unas consideraciones geométricas y pretende capturar los cambios conformacionales, en el sistema químico y en el entorno, esenciales para que la reacción tenga lugar:

$$\Delta\varepsilon = \varepsilon(\boldsymbol{R}) - \varepsilon(\boldsymbol{P}) \qquad (11)$$

Esta coordenada colectiva toma valores negativos en los reactivos, positivos en productos y cercanos a cero en el TS. Este tipo de coordenadas tienen la ventaja adicional de que permiten una introducción más sencilla del efecto túnel ya que no presuponen que el entorno se adapta a los cambios en la posición de la partícula ligera.[48-50]

Formato Deshidrogenasa

En este caso el sistema de estudio es la enzima Formato Deshidrogenasa (FDH) en la que el formiato se reduce a CO2 mediante la transferencia de un hidruro al carbono cofactor NAD (ver Figura 9**¡Error! No se encuentra el origen de la referencia.**). La aplicación de esta coordenada nos permitió el cálculo posterior de la función de onda de los movimientos del hidrógeno trasferido (H41) y del hidrogeno presente en el aceptor (H42) de forma simultánea. Gracias a este tratamiento

pudimos calcular la constante de velocidad de la reacción teniendo en cuenta los efectos quánticos sobre los dos hidrógenos.

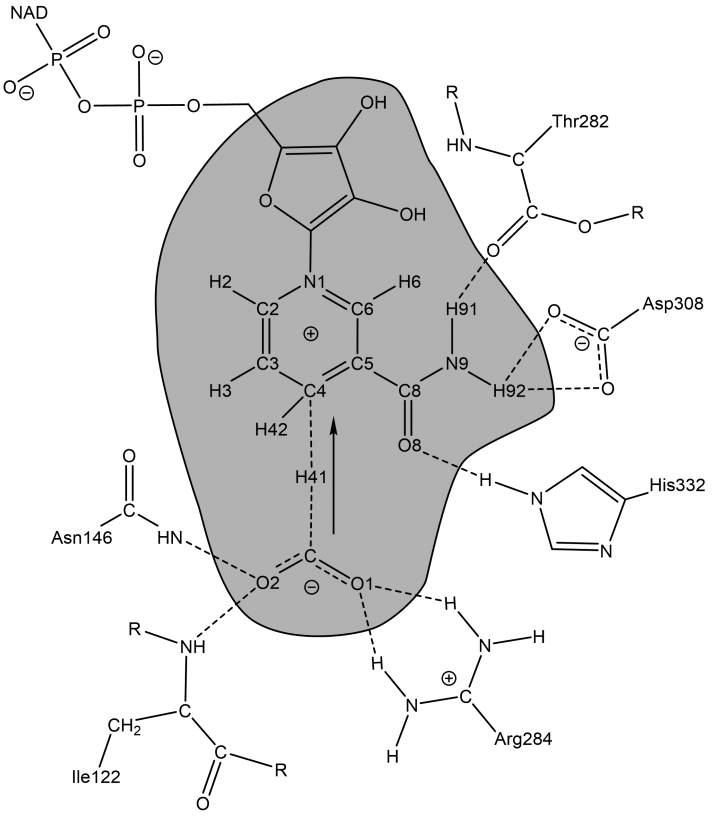


Figura 9: *Representación esquemática del centro activo de la FDH. Las etiquetas de cada átomo se usan como referencia en el texto.*

Para aplicar la coordenada de diferencia energética primero definimos los dos estados en función de las coordenadas de los dos átomos de hidrogeno.

$$\Delta\varepsilon = \varepsilon(\boldsymbol{\zeta}, \boldsymbol{r_1}) - \varepsilon(\boldsymbol{\zeta}, \boldsymbol{r_2}) \qquad (12)$$

Siendo $\boldsymbol{r_1}$ y $\boldsymbol{r_2}$ las coordenadas de los dos átomos de hidrogeno (H41 y H42) en los dos estados definidos (ver Figura 10) y $\boldsymbol{\zeta}$ las coordenadas del resto de átomos.

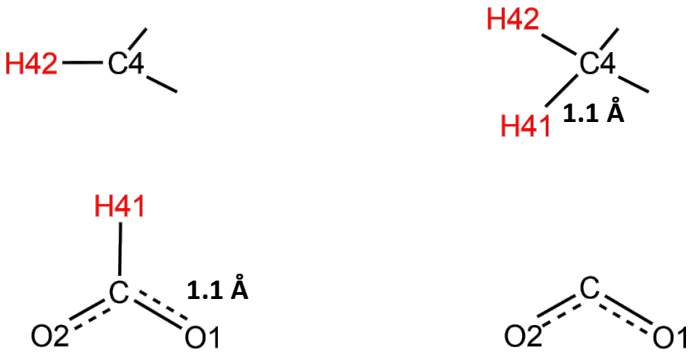


Figura 10: *Representación esquemática de las configuraciones* r_1 *(izquierda) y* r_2 *derecha. Los hidrógenos seleccionados están pintados en rojo.*

La superficie de energía (ver Figura 11) libre se calculó en este sistema en función de la coordenada de diferencia energética ($\Delta\varepsilon$) y de la distancia entre el átomo dador (C) y el átomo aceptor (C4) ya que la transferencia de hidruro es muy sensible a ésta distancia (Q).[49,51]

$$Q = d(CC4) \qquad (13)$$

Los resultados muestran una evolución simultanea de ambas coordenadas hasta llegar a la zona de TS en la que la reacción tiene lugar a lo largo de la coordenada $\Delta\varepsilon$. El TS obtenido para esta coordenada puede presentar una amplio rango de valores de Q, lo que determina la naturaleza de la transferencia de hidruro (adiabático o no adiabático).[49,51]

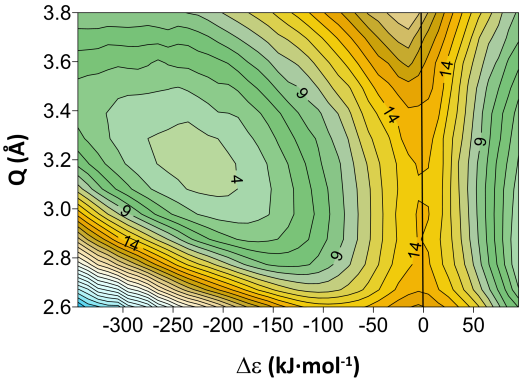


Figura 11: *Superficie de energía libre de la transferencia de hidruro catalizada por la FDH. Las líneas isoenergéticas están dibujadas cada 1* $kcal \cdot mol^{-1}$*. La línea negra representa el conjunto de TS.*

La constante de velocidad debe ser calculada, por lo tanto, como un promedio sobre los diferentes tipos de transferencia que dependen de Q:

$$< k > = \int_0^\infty \rho(Q; \Delta\varepsilon^\ddagger) k_r(Q) dQ \qquad (14)$$

Siendo $\rho(Q; \Delta\varepsilon^\ddagger)$ la distribución de probabilidades de la coordenada Q en el conjunto de TS y $k(Q)$ la constante de velocidad en función de la misma coordenada. La constante de velocidad, incluyendo las propiedades cuánticas de los dos hidrógenos, puede ser expresada de la siguiente manera:[49]

$$k_r(Q) = g(C) \cdot e^{-\frac{\Delta G^\ddagger_{ZPE}(Q)}{RT}} \qquad (15)$$

donde $g(C)$ es un prefactor que depende del acoplamiento (C) entre los estados fundamentales adiabáticos de los dos átomos de hidrógeno en reactivos y productos[49,50] y $\Delta G^\ddagger_{ZPE}(Q)$ es la barrera de energía libre con

una corrección de la energía de punto cero (ZPE) a lo largo de la coordenada Q.

Con el objetivo de analizar la influencia de incluir el efecto túnel, la constante de velocidad será calculada también en el límite adiabático:[49]

$$k_r(Q) = \frac{\omega_s}{2\pi} \cdot e^{-\frac{\Delta G^{\ddagger}_{ZPE}(Q)}{RT}} \tag{16}$$

La distribución de probabilidades $\rho\left(Q; \Delta\varepsilon^{\ddagger}\right)$ se obtuvo a partir de la superficie de energía libre de la Figura 11 mediante la siguiente ecuación:

$$\rho(Q;\ \Delta\varepsilon^{\ddagger}) = \frac{e^{-\frac{\Delta G\left(Q;\ \Delta\varepsilon^{\ddagger}\right)}{RT}}}{\int_0^{\infty} e^{-\frac{\Delta G\left(Q;\ \Delta\varepsilon^{\ddagger}\right)}{RT}} dQ} \tag{17}$$

En este punto puede ser muy ilustrativo comparar la naturaleza de los conjuntos de TS obtenidos usando como coordenada de reacción la coordenada de diferencia energética ($\Delta\varepsilon^{\ddagger}$) o la coordenada de transferencia de hidruro ($q = d(CH41) - d(C4H41)$). Para ello obtuvimos la distribución de probabilidades $\rho\left(Q; q^{\ddagger}\right)$, siendo $q^{\ddagger} = -0.1$ Å, obtenido de un estudio previo sobre el mismo sistema.[52] La Figura 12 muestra que restringir la posición de hidruro para posicionarse en medio de los dos carbonos limita el rango de valores que la coordenada Q puede tomar, cambiando por lo tanto la distancia dador-aceptor a la que la transferencia puede ocurrir.

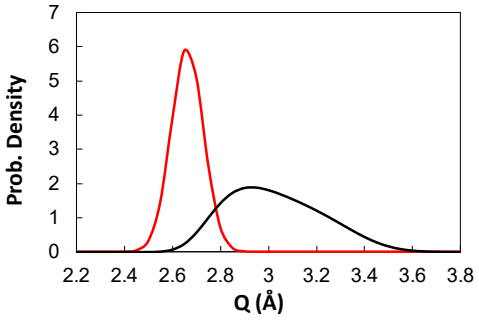


Figura 12: *Distribución de probabilidades de la coordenada Q en el TS definido usando la coordenada Δε (línea negra) y la coordenada q (línea roja).*

Para calcular la corrección de ZPE y el acoplamiento seleccionamos estructuras a diferentes valores de Q de trayectorias calculadas a $\Delta\varepsilon^{\ddagger} = 0\ kJ \cdot mol^{-1}$. Estas estructuras se usaron como coordenadas de partida para obtener las superficies de energía potencial instantánea a lo largo de la coordenada q y el movimiento de flexión del hidrogeno secundario (θ) (ver Figura 13).

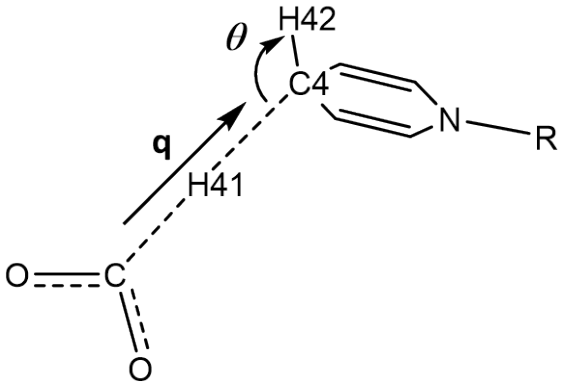


Figura 13: *Representación de las coordenadas usadas las superficies de energía potencial para la transferencia de hidruro en $\Delta\varepsilon^{\ddagger}$.*

Los resultados muestran que la transferencia de hidruro ocurre con un fuerte acoplamiento de ambas coordenadas (ver Figura 14). Además, solo a distancias dador-aceptor mayores de 3.0 Å el nivel vibracional fundamental de los hidrógenos *H*41 y *H*42 está por debajo de la barrera de energía potencial. Esto quiere decir que para distancias menores de 3.0 Å la transferencia es adiabática.

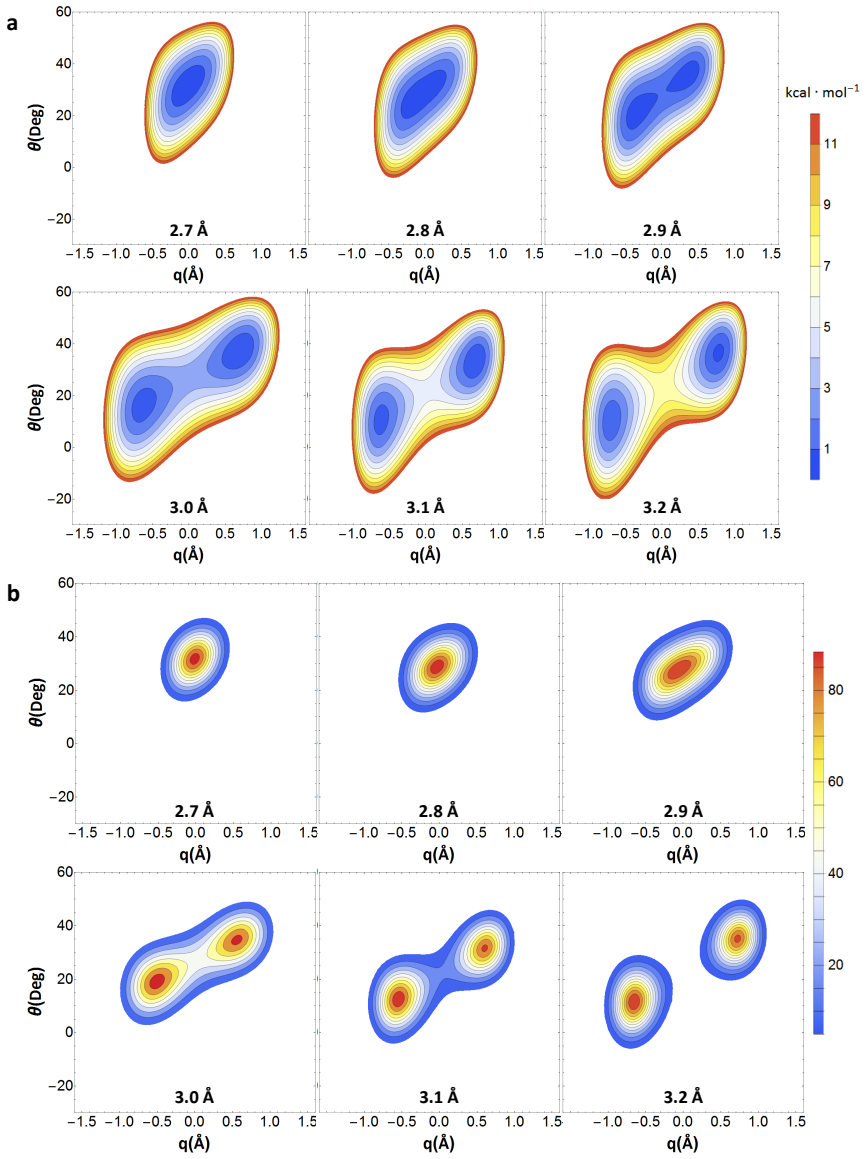


Figura 14: *Superficies de energía potencial promedio a valores de Q 2.7, 2.8, 2.9, 3.0, 3.1 y 3.2 Å como función de la coordenada* q *y el movimiento de flexión del hidrogeno secundario* (a)*. Distribución de probabilidad del primer nivel de la función de onda vibracional calculado a partir de las superficies de energía potencial (*b)

Conocidas la distribución de probabilidades de la distancia dador-aceptor en el TS, la corrección de ZPE y el acoplamiento (C); la constante de velocidad puede ser obtenida usando las ecuaciones 15 y 17. La Figura 15 muestra el valor de la constante de velocidad en función de la coordenada Q a la que la transferencia de hidruro ocurre.

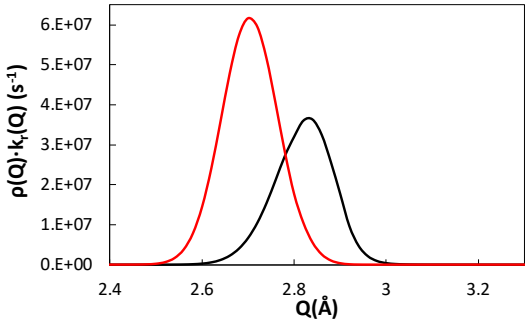


Figura 15: *Perfiles a lo largo de la coordenada Q de la constante de velocidad pesada por la distribución de probabilidad de la distancia dador aceptor para el límite adiabático (línea roja) y el tratamiento completo (negro).*

Estos cálculos, además de estimar la constante de velocidad, nos han permitido obtener los KIEs con unos resultados muy similares a los experimentales.[53,54] Al comparar los resultado obtenidos en el límite adiabático con el tratamiento completo vemos que las diferencias son menores a excepción del KIE secundario, que está subestimado en el límite adiabático (ver Tabla 1). El elevado valor del KIE secundario experimental ha sido atribuido al acoplamiento entre los dos hidrógenos y a la relevancia del túnel durante la transferencia de hidruro en la FDH.[54]

Tabla 1: *Efectos cinéticos isotópicos para el tratamiento completo y para el límite adiabático.*

KIE	Tratamiento completo	Límite adiabático
HH/DH	3.3 ± 0.3	3.3 ± 0.3
HH/HD	1.12 ± 0.04	0.96 ± 0.04
HH/DD	3.4 ± 0.4	3.0 ± 0.4
HH/TH	5.9 ± 0.7	5.7 ± 0.7
DH/DD	1.13 ± 0.05	0.94 ± 0.05
DH/TH	1.8 ± 0.6	1.7 ± 0.6

Los perfiles de la constante de velocidad, con las diferentes sustituciones isotópicas, en función de Q decaen bruscamente al acercarse a distancias de 3.0 Å (ver Figura 16). Es de suponer que un cambio en el perfil de distribución de probabilidades, debido por ejemplo a un incremento de temperatura, no debería afectar significativamente a la relación entre los diferentes perfiles y por tanto los KIEs deberían, en principio, comportarse de forma independiente de la temperatura tal y como se ha observado experimentalmente.[53]

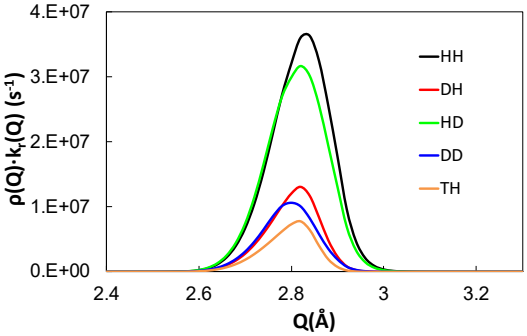


Figura 16: *Perfiles a lo largo de la coordenada Q de la constante de velocidad pesada por la distribución de probabilidad de la distancia dador aceptor para diferentes sustituciones isotópicas.*

Bibliography

(1) Martí, S.; Roca, M.; Andres, J.; Moliner, V.; Silla, E.; Tunon, I.; Bertran, J. *Chem. Soc. Rev.* **2004**, *33*, 98.

(2) Villà, J.; Warshel, A. *J. Phys. Chem. B* **2001**, *105*, 7887.

(3) García-Viloca, M.; Poulsen, T. D.; Truhlar, D. G.; Gao, J. *Protein Sci.* **2004**, *13*, 2341.

(4) Marti, S.; Andres, J.; Moliner, V.; Silla, E.; Tunon, I.; Bertran, J. *Chem. Soc. Rev.* **2008**, *37*, 2634.

(5) Modak, J. M. *Reson* **2002**, *7*, 69.

(6) Lee, C. C.; R., M. W.; Hu, Y. In *The Metal-Driven Biogeochemistry of Gaseous Compounds in the Environment*; Kroneck, P. M. H., Torres, M. E. S., Eds.; Springer Netherlands: 2014; Vol. 14, p 147.

(7) McGuire, J. J. *Curr. Pharm. Des.* **2003**, *9*, 2593.

(8) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.

(9) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700.

(10) Henzler-Wildman, K. A.; Lei, M.; Thai, V.; Kerns, S. J.; Karplus, M.; Kern, D. *Nature* **2007**, *450*, 913.

(11) Hammes, G. G.; Benkovic, S. J.; Hammes-Schiffer, S. *Biochemistry* **2011**, *50*, 10422.

(12) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. *Chem. Rev.* **2006**, *106*, 3210.

(13) Karplus, M.; McCammon, J. A. *Ann. Rev. Biochem.* **1983**, *52*, 263.

(14) García-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Science* **2004**, *303*, 186.

(15) Kamerlin, S. C. L.; Warshel, A. *Proteins* **2010**, *78*, 1339.

(16) Antoniou, D.; Basner, J.; Núñez, S.; Schwartz, S. D. *Chem. Rev.* **2006**, *106*, 3170.

(17) Gao, J.; Ma, S.; Major, D. T.; Nam, K.; Pu, J.; Truhlar, D. G. *Chem. Rev.* **2006**, *106*, 3188.

(18) Olsson, M. H. M.; Parson, W. W.; Warshel, A. *Chem. Rev.* **2006**, *106*, 1737.

(19) Nashine, V. C.; Hammes-Schiffer, S.; Benkovic, S. J. *Curr. Opin. Chem. Biol.* **2010**, *14*, 644.

(20) Adamczyk, A. J.; Cao, J.; Kamerlin, S. C. L.; Warshel, A. *Proc. Nat. Acad. Sci. USA* **2011**, *108*, 14115.

(21) Bhabha, G.; Lee, J.; Ekiert, D. C.; Gam, J.; Wilson, I. A.; Dyson, H. J.; Benkovic, S. J.; Wright, P. E. *Science* **2011**, *332*, 234.

(22) Ramanathan, A.; Agarwal, P. K. *PLOS Biol.* **2011**, *9*, e1001193.

(23) Zhang, J.; Klinman, J. P. *J. Am. Chem. Soc.* **2011**, *133*, 17134.

(24) Glowacki, D. R.; Harvey, J. N.; Mulholland, A. J. *Nat. Chem.* **2012**, *4*, 169.

(25) García-Meseguer, R.; Martí, S.; Ruiz-Pernía, J. J.; Moliner, V.; Tuñón, I. *Nat. Chem.* **2013**, *5*, 566.

(26) Ruiz-Pernía, J. J.; Luk, L. Y. P.; García-Meseguer, R.; Martí, S.; Loveridge, E. J.; Tuñón, I.; Moliner, V.; Allemann, R. K. *J. Am. Chem. Soc.* **2013**, *135*, 18689.

(27) García-Meseguer, R.; Zinovjev, K.; Roca, M.; Ruiz-Pernía, J. J.; Tuñón, I. *J. Phys. Chem. B* **2015**, *119*, 873.

(28) Wigner, E. *Trans. Far. Soc.* **1938**, *34*, 29.

(29) Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J. *J. Phys. Chem.* **1996**, *100*, 12771.

(30) Tuñón, I.; Laage, D.; Hynes, J. T. *Arch. Biochem. Biophys.* **2015**, *582*, 42.

(31) Pu, J.; Gao, J.; Truhlar, D. G. *Chem. Rev.* **2006**, *106*, 3140.

(32) Bergsma, J. P.; Gertner, B. J.; Wilson, K. R.; Hynes, J. T. *J. Chem. Phys.* **1987**, *86*, 1356.

(33) Stucki, G.; Thueer, M. *Environ. Sci. Technol.* **1995**, *29*, 2339.

(34) Janssen, D. B.; Scheper, A.; Dijkhuizen, L.; Witholt, B. *Appl. Environ. Microbiol.* **1985**, *49*, 673.

(35) Vidgren, J.; Svensson, L. A.; Liljas, A. *Nature* **1994**, *368*, 354.

(36) Schlissel, M. S.; Brown, D. D. *Cell* **1984**, *37*, 903.

(37) Thoma, F.; Koller, T.; Klug, A. *J. Cell. Biol.* **1979**, *83*, 403.

(38) Sawaya, M. R.; Kraut, J. *Biochemistry* **1997**, *36*, 586.

(39) Osborne, M. J.; Schnell, J.; Benkovic, S. J.; Dyson, H. J.; Wright, P. E. *Biochemistry* **2001**, *40*, 9846.

(40) Arora, K.; Brooks III, C. L. *J. Am. Chem. Soc.* **2009**, *131*, 5642.

(41) Boehr, D. D.; McElheny, D.; Dyson, H. J.; Wright, P. E. *Science* **2006**, *313*, 1638.

(42) Li, L.; Wright, P. E.; Benkovic, S. J.; Falzone, C. J. *Biochemistry* **1992**, *31*, 7826.

(43) Cameron, C. E.; Benkovic, S. J. *Biochemistry* **1997**, *36*, 15792.

(44) Miller, G. P.; Benkovic, S. J. *Biochemistry* **1998**, *37*, 6336.

(45) Benkovic, S. J.; Hammes-Schiffer, S. *Science* **2003**, *301*, 1196.

(46) Loveridge, E. J.; Behiry, E. M.; Guo, J.; Allemann, R. K. *Nat. Chem.* **2012**, *4*, 292.

(47) Luk, L. Y. P.; Ruiz-Pernía, J. J.; Dawson, W. M.; Roca, M.; Loveridge, E. J.; Glowacki, D. R.; Harvey, J. N.; Mulholland, A. J.; Tuñón, I.; Moliner, V.; Allemann, R. K. *Proc. Nat. Acad. Sci. USA* **2013**, *110*, 16344.

(48) Borgis, D.; Hynes, J. T. *Chem. Phys.* **1993**, *170*, 315.

(49) Borgis, D. C.; Hynes, J. T. *J. Phys. Chem.* **1996**, *100*, 1118.

(50) Kiefer, P. M.; Hynes, J. T. *J. Phys. Org. Chem.* **2010**, *23*, 632.

(51) Borgis, D. C.; Lee, S.; Hynes, J. T. *Chem. Phys. Lett.* **1989**, *162*, 19.
(52) Castillo, R.; Oliva, M.; Martí, S.; Moliner, V. *J. Phys. Chem. B* **2008**, *112*, 10012.
(53) Bandaria, J. N.; Cheatum, C. M.; Kohen, A. *J. Am. Chem. Soc.* **2009**, *131*, 10151.
(54) Hermes, J. D.; Morrical, S. W.; O'Leary, M. H.; Cleland, W. W. *Biochemistry* **1984**, *23*, 5479.